Université de Sherbrooke

Le viroïde de la mosaïque latente du pêcher (PLMVd): un modèle pour mieux comprendre l'évolution et la biologie des viroïdes du groupe A

Par

M. Frédéric Bussière

Département de Biochimie

Thèse présentée à la Faculté de Médecine

en vue de l'obtention du grade de

philosophiae doctor (Ph.D.) en Biochimie

8 décembre 1999

TABLE DES MATIERES

LISTE DES FIGURES

INTRODUCTION

RÉSULTATS

1.1 ARTICLE

DISCUSSION

LISTE DES TABLEAUX

RÉSULTATS

1.1 ARTICLE

2.1 ARTICLE

3.1 ARTICLE

4.1 ARTICLE

4.2 ARTICLE

5.1 ARTICLE

LISTE DES SYMBOLE ET ABRÉVIATIONS

A	adénosine
ARN	acide ribonucléique
ARNm	ARN messager
ARN pol I	ARN polymérase I
ARN pol II	ARN polymérase II
ARN pol III	ARN polymérase III
ARNr	ARN ribosomal
ASBVd	*avocado sunblotch viroid*
C	cytosine
CarSV	*carnation small viroid-like RNA*
CbVd1	*coleus blumei viroid*
CCCVd	*coconut cadang-cadang viroid*
CChMVd	*chrysanthemum chlorotic mottle viroid*
CERV	*carnation etched ring caulimovirus*
DIPA	*human hepatitis delta antigen interacting protein A*
G	guanosine
Kcat	constante catalytique
kDa	kilodalton
Licl	chlorure de lithium
M	moles par litre
Mg^{2+}	magnésium
nt	nucléotide(s)
pb	paire(s) de bases
PCR	réaction de polymérisation en chaine
PLMVd	*peach latent mosaic viroid*
pré-ARNm	ARN pré-messager

PSTVd	*potato spindle tuber viroid*
RdRp	*RNA dependant RNA polymerase*
rNS2T	*newt satellite 2 RNA transcript*
RT-PCR	*reverse-transcribed*-PCR
S	Svedberg
sARMV	*arabis mosaic virus satellite RNA*
sCYMV	*chicory yellow mottle virus satellite RNA*
sn RNA	petit ARN nucléaire
sTobRV	*tobacco ringspot virus satellite RNA*
U	uracile
vBYDV	*barley yellow dwarf virus satellite RNA*
(+)	identifie la polarité positive, ou brin plus de l'ARN
(-)	identifie la polarité négative, ou brin moins de l'ARN

RÉSUMÉ

Le viroïde de la mosaïque latente du pêcher (PLMVd): un modèle pour mieux comprendre l'évolution et la biologie des viroïdes du groupe A.

Par M. Frédéric Bussière

Département de Biochimie

Université de Sherbrooke

Thèse présentée à la Faculté de Médecine

en vue de l'obtention du grade de

philosophiae doctor (Ph.D.) en Biochimie

8 décembre 1999

Les viroïdes sont de courts ARN simple brin circulaires (~300 nucléotides) qui infectent les plantes supérieures causant des pertes économiques importantes en agriculture. Ils se répliquent à l'intérieur des cellules infectées par un mécanisme en cercle roulant (Branch et Robertson, 1984). Leur génome ne code apparemment pour aucune protéine et ils sont ainsi dépendant de leur hôte pour assurer certaines étapes de leur cycle vital. On connaît aujourd'hui 29 espèces de viroïdes appartenant à deux groupes distincts. Les viroïdes du groupe B possèdent certains éléments de séquence et de structure qui sont conservés. Leur génome est subdivisé en 5 domaines caractéristiques auxquels on attribue certaines fonctions biologiques (Keese et Symons, 1985; Flores et al., 1997). Ces ARN se retrouvent au niveau du noyau des cellules infectées où ils seraient répliqués par l'ARN polymérase II. Les viroïdes du groupe A ne possèdent pas l'organisation génomique typique des viroïdes du groupe B et on ne dénote aucune conservation au niveau de leur séquence ou de leur structure. Ils sont caractérisés par la

présence de motifs autocatalytiques (*hammerhead*) qui permettent la coupure des brins multimériques produits lors de la réplication en cercle roulant. Ces motifs autocatalytiques sont également présents chez certains ARN satellites de virus de plantes. Ces ARN diffèrent des viroïdes du groupe A puisqu'ils sont dépendant de la présence d'un virus associé pour leur réplication et leur propagation. Les nombreuses similarités entre ces ARN satellites et les viroïdes suggèrent une origine évolutive commune (Branch et al., 1990).

La méconnaissance de la biologie des viroïdes du groupe A est en grande partie attribuable au peu d'ampleur des recherches entreprises sur ces agents pathogènes. En 1994, on ne connaissait que deux espèces appartenant à ce groupe, soit ASBVd et PLMVd. Nous avons donc initié des études visant à mieux comprendre certains aspects de leur biologie moléculaire, tels le cycle de réplication et la localisation cellulaire. Nous avons utilisé PLMVd comme modèle d'étude. Cet ARN de 338 nucléotides possède des motifs *hammerhead* qui assurent l'autocoupure des brins multimériques (Hernandez et Flores, 1992). La caractérisation *in vitro* des mécanismes régissant l'autocoupure et l'autoligation (Côté et Perreault, 1997) permet d'interpréter les particularités du cycle de réplication de PLMVd observées *in vivo*. Les résultats obtenus confirment également la localisation chloroplastique des viroïdes du groupe A (Lima et al., 1994). Cette différence fondamentale avec les membres du groupe B explique probablement l'importante divergence entre ces deux groupes de viroïdes. Elle a donc des implications majeures au niveau de leur évolution et de leur biologie. Ces résultats suggèrent également qu'une ARN polymérase chloroplastique est probablement impliquée dans la réplication des viroïdes du groupe A. Finalement, les études structurales révèlent l'existence de plusieurs éléments qui semblent conservés.chez certains viroïdes du groupe A. Ces résultats sont un premier pas vers la compréhension des relations structure/fonction chez ces ARN.

Parallèlement, nous avons mis sur pied une banque de données contenant les séquences et les informations pertinentes pour tous les viroïdes et ARN apparentés connus à ce jour (http://www.callisto.si.usherb.ca/~jpperra). Nous avons également proposé une nouvelle nomenclature pour pallier la confusion croissante qui accompagne la caractérisation de nouvelles séquences par les différents groupes de recherche. Cette ressource a été utilisée afin de tenter la réalisation d'études phylogénétiques. La grande divergence de la séquence de leur génome est cependant un obstacle majeur à l'étude de leurs relations évolutives. Le motif catalytique *hammerhead* semble être le seul élément conservé entre les viroïdes du groupe A et les ARN satellites. Une étude plus approfondie des *hammerhead* supporte une origine évolutive commune de ces ARN. Nous avons donc utilisé ces régions comme un caractère homologue afin de reconstituer leur histoire évolutive. Le modèle proposé suggère que les viroïdes aient évolué à partir d'une forme ancestrale des ARN satellites. On ignore cependant l'origine de ces ARN. L'existence de motifs *hammerhead* dans certains transcrits d'ADN répétitif suggère une apparition à partir d'éléments présents dans le génome cellulaire (Epstein et Gall, 1987; Ferbeyre et al., 1998). L'identification éventuelle d'autres *hammerhead* dans les banques de données et leur caractérisation promet d'apporter des informations intéressantes afin de mieux comprendre l'origine et l'évolution de ces agents pathogènes des plantes.

Mots-clés: viroïdes, *hammerhead*, ribozyme, ARN catalytique

INTRODUCTION

1. LES VIROIDES

1.1 Généralités

À la fin des années 1960 quelques laboratoires tentaient d'identifier l'agent étiologique d'une maladie très répandue qui affligeait la pomme de terre et causait ainsi de lourds degâts dans les champs cultivés. De toute évidence, il ne s'agissait pas d'un virus conventionnel. Deux maladies causées par le même type d'agent pathogène, et affectant les citronniers et les chrysanthèmes, furent également identifiées à cette époque. En 1967, le groupe du Dr. Diener démontra que la maladie de la pomme de terre était en fait causée par un acide nucléique libre (non encapsidé) (Diener et Raymer, 1967). Il s'agissait d'un ARN de faible poids moléculaire qui se répliquait à l'intérieur des cellules infectées (Diener, 1971). Ces ARN pathogènes reçurent alors l'appellation de viroïdes afin de les distinguer des infections virales conventionnelles. De nombreuses maladies des plantes attribuables aux viroïdes furent par la suite identifiées. Plusieurs d'entre elles causent des pertes économiques considérables. À titre d'exemple, sur l'île de San Miguel, dans les Philippines, la presque totalité des 250 000 palmiers sont morts suite à l'infection par CCCVd en 1962 (Haseloff et al., 1982).

1.1.1 Réplication des viroïdes

Les viroïdes sont de courts ARN simple brin circulaires non codant qui infectent les plantes vasculaires. Le génome des viroïdes est non encapsidé et sa taille varie de 246 à 475 nt,

selon les espèces. On dénombre actuellement 29 espèces de viroïdes qui appartiennent à deux groupes ayant des caractéristiques bien distinctes (groupes A et B, voir section 1.3). Ces ARN non-codants dépendent des protéines et systèmes enzymatiques de la cellule hôte qui assurent certaines activités catalytiques ou structurales essentielles (e.g. réplication). Tous ces ARN sont répliqués à l'intérieur des cellules hôtes par un mécanisme en cercle roulant (fig. 1) (Branch et Robertson, 1984). Le brin circulaire infectieux, auquel on attribue arbitrairement la polarité positive ((+)), est utilisé comme matrice pour la synthèse d'un brin multimérique complémentaire (polarité négative, (-)). Celui-ci est ensuite coupé de façon à produire des copies monomériques qui sont circularisées pour donner des ARN circulaires négatifs (-). Pour compléter le cycle, les même étapes de polymérisation, coupure et ligation sont répétées en utilisant cette fois l'intermédiare circulaire (-) comme matrice. Ce cycle est dit symmétrique puisque la biosynthèse et la maturation des intermédiaires de polarité (+) et (-) impliquent des étapes identiques. Par comparaison, durant un cycle de réplication asymmétrique, l'intermédiaire multimérique (-) est directement utilisé pour la synthèse du brin multimérique (+) (voir fig. 1). Sa coupure et la ligation des monomères ainsi produits viennent compléter la réplication. Il n'y a donc pas production d'ARN circulaire (-). Les viroïdes sont répliqués par un cycle symmétrique ou asymmétrique, dépendamment des espèces (voir section 1.3).

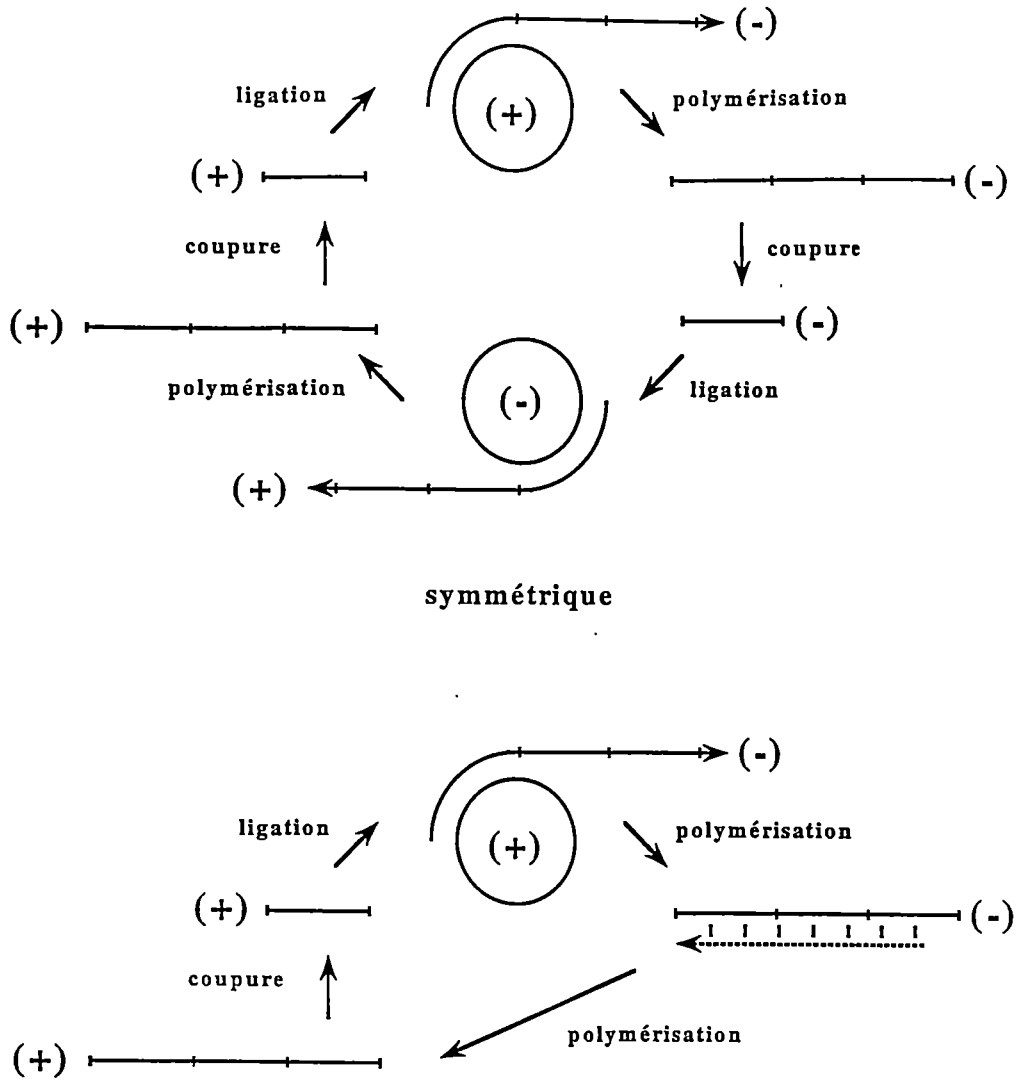


Figure 1. Réplication en cercle roulant. L'ARN circulaire infectieux, auquel on attribue arbitrairement la polarité positive (ou plus, (+)), est utilisé comme matrice pour la synthèse d'un ARN multimérique complémentaire négatif (ou moins, (-)). Sa coupure et la ligation des monomères ainsi produits permettent la formation des intermédiaires circulaires (-). Dans le cycle symmétrique les même étapes de polymérisation , coupure et ligation, en utilisant la forme circulaire (-) comme matrice, permettent de compléter la réplication. Dans le cycle asymmétrique, l'ARN multimérique (-) est directement utilisé pour la synthèse de l'ARN multimérique (+) qui est ensuite coupé et circularisé.

1.1.2 Transmission et pathogénèse

Les viroïdes sont principalement transmissibles par brisure mécanique; mécanisme répandu chez les virus qui parasitent les végétaux. L'agent infectieux est transmis suite à un contact, au niveau d'une blessure, entre un plant infecté (ou un outil infecté) et un plant non-infecté. Suite à son entré dans une cellule hôte, le viroïde peut s'y répliquer et se propager à l'intérieur de l'organisme en utilisant vraisembablement les jonctions cellulaires (plasmodesmes) pour passer d'une cellule à l'autre (Ding et al., 1997). À l'instar des infections virales chez les plantes, l'infection par un viroïde peut être soit asymptomatique ou symptomatique. Une infection symptomatique peut également être latente pour une certaine période ou avoir des effets immédiats. Finalement, on observe une importante gamme de symptômes, pouvant varier d'une simple décoloration du feuillage à la mort complète du plant infecté, en passant par le rabougrissement et la nécrose des tissus. Une même espèce de viroïde peut également provoquer des symptômes et un degré de sévérité de la maladie qui sont variables. De telles différences peuvent parfois être observées chez deux variants (ou souches) dont la séquence du génome ne diffère que par quelques nucléotides. Comme pour les virus qui infectent les plantes, très peu de choses sont connues sur les mécanismes de pathogénécité des viroïdes. Chez les membres du groupe B, une région de leur génome (domaine de pathogénécité) semble associée à la pathogénécité puisqu'on y retrouve la plupart des mutations qui affectent le type et la sévérité des symptômes. Cependant, d'autres domaines semblent également intervenir (voir section 1.3.1).

Plusieurs mécanismes ont été proposés pour tenter d'expliquer comment les viroïdes peuvent exercer leur effet pathogène. Il existe cependant peu d'évidences expérimentales pour appuyer les nombreuses hypothèses émises. Étant donné que ces ARN ne semblent coder pour aucun polypeptide, leur effet serait vraisemblablement médié par des interactions directes avec des composantes cellulaires de l'hôte (revue par Diener et al., 1993b). Les

prochains paragraphes résument brièvement les candidats potentiels ainsi que les hypothèses ayant vu le jour.

Interactions viroïdes-protéines. On recherche depuis de nombreuses années des protéines cellulaires capable d'interagir de façon spécifique avec les viroïdes afin d'élucider les différents mécanismes impliqués dans leur cycle biologique (e.g. réplication, pathogénèse). On a, par exemple, démontré que PSTVd était retrouvé principalement sous forme de particules ribonucléoprotéiques, associé aux nucléosomes. PSTVd interagirait de façon stable non seulement avec les histones qui composent les nucléosomes, mais également avec une protéine nucléaire inconnue de 41 kDa (Wolff et al., 1985). De plus, les expériences menées dans notre laboratoire suggèrent que PLMVd, un viroïde du groupe A, interagirait de façon spécifique avec des protéines de 23 kDa et 10 kDa (Bussière et Perreault, résultats non publiés). Il n'existe cependant aucune évidence démontrant le rôle de ces interactions avec le processus de pathogénèse.

Interactions viroïdes-ARN cellulaires. Certains viroïdes possèdent des régions qui sont complémentaires à des ARN cellulaires. Ces observations suggèrent que la pathogénécité pourrait être le résultat d'un effet antisens bloquant l'expression (ARN codants) ou la fonction (ARN structuraux) de certains gènes. On a par exemple dénoté une complémentarité de séquence entre une partie du brin (-) de PSTVd et l'ARN U1 de mammifère (snRNA U1) (Diener, 1981). PSTVd pourrait ainsi nuire à la maturation des pré-ARNm. Le groupe de Sanger a également caractérisé l'existence de régions complémentaires entre cinq expèces de viroïdes qui infectent les plants de tomates et l'ARN 7S de tomates, qui fait partie de la particule de reconnaissance du signal (Haas et al., 1988). L'infection pourrait donc interférer dans le processus de translocation cotraductionnelle de certaines protéines à l'intérieur du réticulum endoplasmique. Bien qu'à prime-à-bord attrayantes, ces hypothèses souffrent de l'absence de preuves expérimentales

et ne permettent pas d'expliquer pourquoi deux variants d'une même espèce de viroïde, dont la séquence ne diffère que par quelques nucléotides, causent l'apparition de symptômes de nature et de sévérité différentes. De plus, les séquences impliquées dans la pathogénèse sont assez variables et les mutations observées sont souvent retrouvées à l'intérieur des régions de complémentarité. À la lumière des récentes connaissances acquises sur les régions régulatrices de la pathogénécité (voir section 1.3.1), il semble plus probable d'envisager l'implication d'éléments structuraux dans ce phénomène.

Deux autres hypothèses ont été finalement émises afin d'expliquer l'induction de maladies par les viroïdes. La première, le mimétisme moléculaire, suppose que par des homologies de séquences et/ou de structures un viroïde pourrait remplacer un ARN cellulaire dans sa fonction et ainsi désorganiser les processus moléculaires qui en découlent. De telles homologies ont été rapportées entre PSTVd, les introns du groupe I et l'ARN U3B (snRNA U3B) (Dinter-Gottlieb, 1986; Kiss et al., 1983). La deuxième hypothèse concerne les viroïdes du groupe A (voir section 1.3.2) qui possèdent des motifs catalytiques et pourraient ainsi médier la coupure spécifique d'ARN cellulaires. Des études dans notre laboratoire sont actuellement en cours afin de vérifier la validité de cette dernière hypothèse.

Jusqu'à maintenant il existe donc peu de résultats tangibles en ce qui concerne l'identification des composantes de l'hôte qui interagissent de façon spécifique avec le génome des viroïdes et déclenchent les processus de pathogénécité. On a cependant identifié des protéines qui sont activées ou surexprimés en réponse à l'infection par un viroïde. Par exemple l'expression des protéines PR-1 et PR-69 est considérablement stimulée suite à l'infection chez *Lycopersicon esculentum* (tomate) (Tornero et al., 1994 et 1996). PR-1 et PR-69 font partie de la famille des protéines PR (Pathogenesis-Related proteins) qui est composée de gènes impliqués dans la réponse de défense aux agents pathogènes. Chez *Lycopersicon esculentum* on a également démontré qu'une protéine de 68 kDa était

phosporylée suite à l'infection par PSTVd. La protéine est immunologiquement homologue à la protéine kinase P68 humaine qui est activée (autophosphorylée) par l'ARN double-brin (Hiddinga et al., 1988). Cette protéine kinase de tomate peut être activée *in vitro* par l'addition du viroïde et son niveau de phosphorylation semble être relié au type de variant utilisé. Par exemple une souche sévère de PSTVd provoque un plus haut niveau de phosphorylation de P68 qu'une souche intermédiaire (Diener et al., 1993a). Cette corrélation laisse croire que P68 est un candidat intéressant pour la compréhension des mécanismes qui initient la pathogénécité.

1.2 Les viroïdes versus les ARN satellites apparentés

Certaines infections virales sont accompagnées d'une coinfection par un ARN de petite taille, appelé ARN satellite. Celui-ci ne possède aucune homologie de séquence avec le génome viral (revue par Roossinck et al., 1992). Son rôle sur l'induction des symptômes chez les végétaux dépend de trois facteurs: la souche virale, la nature de l'ARN satellite et celle du plant hôte. Bien que dans la plupart des cas la présence d'un ARN satellite semble atténuer les symptômes dûs à l'infection virale, une augmentation de la sévérité de la maladie peut également survenir. Les ARN satellites sont des ARN infectieux qui sont dépendants de la présence d'un virus à ARN (virus associé) pour leur viabilité. En effet, le génome du virus associé encode les protéines structurales nécessaires à sa propre encapsidation ainsi qu'à celle de l'ARN satellite. Cette coencapsidation permet la transmission simultanée de l'ARN satellite et du virus associé à des organismes non infectés, assurant ainsi cette association indispensable, pour l'ARN satellite, à l'intérieur du nouvel hôte. De plus, le virus associé est responsable de la multiplication de l'ARN satellite puisque son génome encode également l'ARN polymérase nécessaire à leur réplication. Les ARN satellites sont donc des parasites moléculaires opportunistes des systèmes viraux, autant chez les plantes que chez les animaux. La taille du génome des ARN satellites peut

varier de 200 à 1700 nt, environ. Les analyses de séquences ont permis d'identifier des cadres de lecture ouverts potentiels qui pourraient être exprimés *in vivo* et ainsi participer à la réplication, au transport ainsi qu'à la pathogénèse de ces ARN. Certaines observations suggèrent que ces cadres de lecture ouverts seraient fonctionnels *in vivo*. Par exemple, en dépit d'une variation importante de la séquence entre plusieurs isolats d'un même ARN satellite, on retrouve d'importantes homologies au niveau de la séquence en acides aminés. Les approches expérimentales ainsi que les analyses de séquences suggèrent cependant que cette faculté ne serait associée qu'à certains ARN satellite dont la taille du génome est supérieure à environ 700 nt. Parmis les ARN satellites de virus de plantes dont la taille est inférieure à environ 600 nt, on retrouve des espèces qui accumulent un intermédiaire de réplication sous forme circulaire et possèdent des motifs catalytiques tels que ceux retrouvés chez trois espèces de viroïdes (viroïdes du groupe A, voir section 1.3.2 et section 2). Comme pour les viroïdes, le génome de ces ARN satellites est fortement structuré et ne coderait pour aucune protéine. On considère donc ces ARN comme étant apparentés. En raison de l'accumulation prépondérante d'une forme circulaire qui rappelle les viroïdes, certains sont même désignés sous le terme virusoïdes dans la littérature scientifique.

On dit des viroïdes qu'ils sont autonomes par rapport à la dépendance des ARN satellites face à la présence d'un virus associé. Les viroïde utilisent les systèmes enzymatiques de la cellule hôte pour les différentes étapes de leur cycle vital (i.e. leur réplication, leur transmission ainsi que leur propagation à l'intérieur de l'organisme). Certains travaux ont cependant montré que PSTVd serait encapsidé et transmis par le virus PLRV (potato leaf roll virus) (Querci et al., 1997). Les viroïdes pourraint donc être transmis également par d'autres mécanismes que la brisure mécanique lorsque coencapsidés avec un génome viral. Cette association, bien qu'elle ne soit pas indispensable au viroïde, rappelle celle observée chez les ARN satellites. Elle pourrait permettre des modes de transmission plus variées. On connaît en effet certains exemples de propagation par des insectes suceurs

chez les viroïdes (Querci et al., 1997; Desvignes, 1986). Par exemple, le viroïde de la mosaïque latente du pêcher (PLMVd) peut être transmis par *Myzus persicae* Sulzer (Desvignes, 1986).

1.3 Classification

1.3.1 Viroïdes du groupe B

Les viroïdes du groupe B sont de loin ceux ayant fait l'objet des recherches les plus intensives. Par conséquent, des 29 espèces de viroïdes connues, 26 appartiennent au groupe B. De plus, la majeure partie de nos connaissances acquises sur la biologie des viroïdes concerne plus particulièrement les espèces appartenant à ce groupe. Les prédictions de structures secondaires par ordinateur suggèrent que ces ARN formeraient une longue tige principalement double brin (fig. 2). Leur génome est divisé en 5 domaines bien définis dont la séquence est plus ou moins conservée entre les espèces (fig. 2) (Keese et Symons, 1985). Le domaine C (*conserved*) ou CCR (*conserved central region*), qui comprend environ 15 pb, existe sous deux formes différentes, ce qui a mené à une subdivision du groupe B en deux sous-groupes, B1 et B2. Le CCR est un domaine très peu variable et est de loin le plus conservé pour chacun des sous-groupe. La caractérisation en 1990 de CbVd1 suggère l'existence d'un troisième sous-groupe puisque son CCR diffère de celui des deux sous groupes B1 et B2 (Spieker et al., 1990). Le CCR contient une structure tertiaire appelé *loop-E* (Paul et al., 1992) qui semble jouer un rôle important dans le repliement de nombreux ARN (Leontis et Westhof, 1998) (fi.g 2). Il peut également adopter des structures alternatives en forme de courtes tige-boucles (e.g. *hairpin I* (hpI)). Le CCR serait responsable des évènements de maturation des brins multimériques produits lors de la réplication (coupure et ligation, voir fig. 1). Les domaine P (*pathogenecity*) et TL (*terminal left*) sont, quant à eux, impliqués dans le phénomène de pathogénécité (Sano et al., 1992).

On a par exemple observé une corrélation entre la stabilité d'une partie de la région P (dans la structure native double brin) et la sévérité des symptômes. Finalement, les régions V (variable) et TR (*terminal right*) joueraient un rôle dans l'accumulation du viroïde dans les cellules infectées en agissant au niveau de la réplication, de la stabilité et/ou du transport de l'ARN. Ces domaines sont également impliqués dans l'induction des maladies puisqu'un lien direct entre la vitesse d'apparition des symptômes et celle de l'accumulation du viroïde a été observé (Sano et Ishiguro, 1998). À l'intérieur du domaine TL il existe une région appelée TCR (*terminal conserved region*) qui est présente chez tous les viroïdes du groupe B dont la taille est supérieure à environ 300 nt (20 espèces sur 29). À la différence du CCR, la séquence de cet élément est très conservée entre les sous-groupe B1 et B2. Cette caractéristique suggère une fonction importante et est considérée comme candidat de choix pour le recrutement des facteurs nécessaires à la réplication des viroïdes du groupe B (Flores et al., 1997).

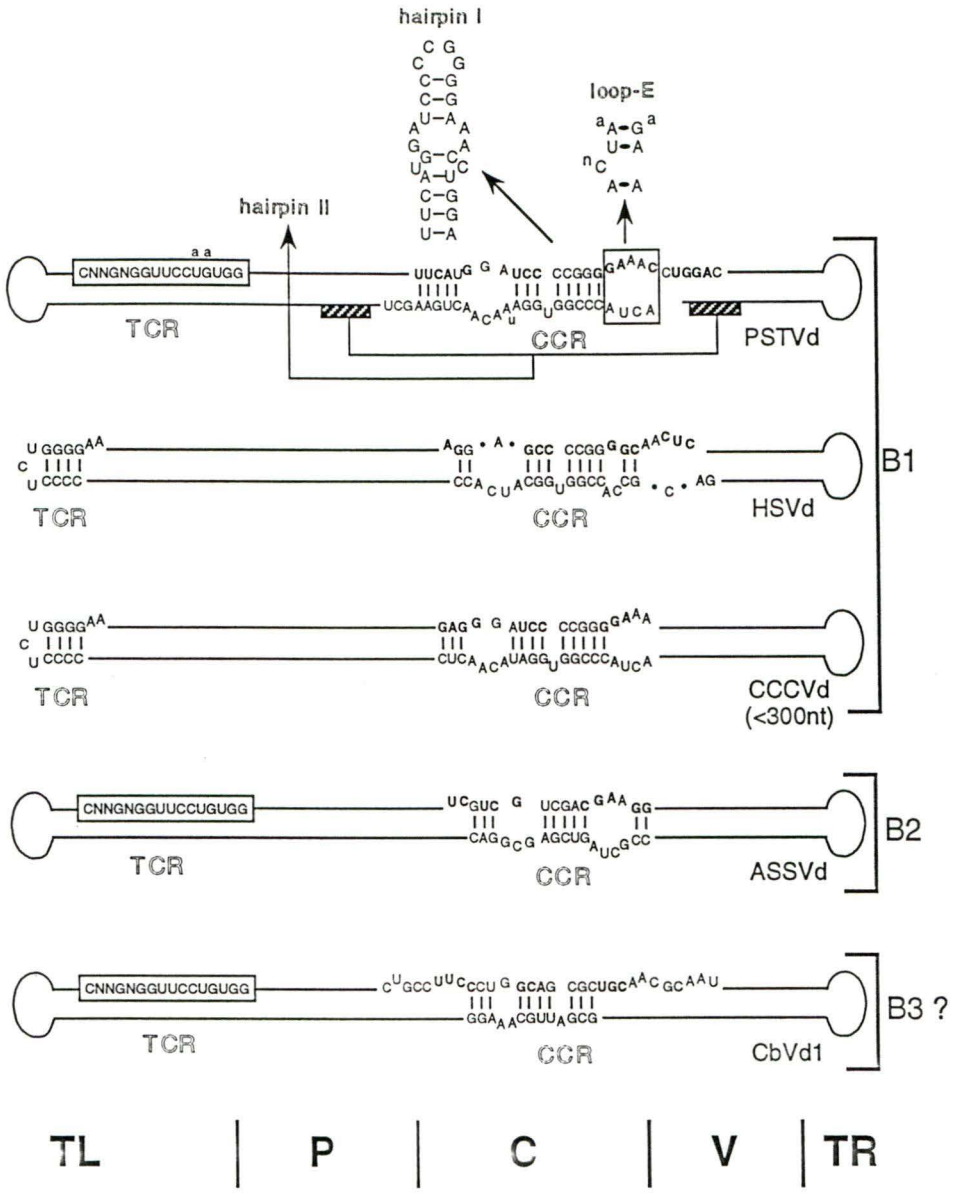


Figure 2. L'organisation des 5 domaines génomiques chez les viroïdes du groupe B. Représentation schématique de l'organisation des domaines conservés pour les sous-groupes B1 et B2: TL (*terminal left*), P (*pathogenecity*), C (*conserved*), V (variable), TR (*terminal right*), CCR (*central conserved region*) et TCR (*terminal conserved region*) (voir section 1.3.1). CbVd1 pourrait appartenir à un troisième sous-groupe (B3). La structure la plus stable de la région CCR d'un représentant de chaque sous-groupe est montrée. Les structures secondaires et tertiaires caractérisées à l'intérieur de la région centrale de ces ARN sont identifiées sur le génome de PSTVd (*hairpin I* et *loop-E*). La structure en épingle à cheveux *hairpin II* bordant le CCR est également identifiée.

On considère généralement l'ARN polymérase II comme étant responsable de la réplication des viroïdes du groupe B. Les études d'inhibition par l'α-amanitine (inhibiteur de l'ARN pol II) et la localisation cellulaire de certains des viroïdes du groupe B au niveau du noyau des cellules infectées demeurent parmi les principales évidences de son implication (Mühlbach et Sänger, 1979; Semancik et Harper, 1984; Harders et al., 1989; Semancik et al., 1976). On s'interroge toutefois encore sur l'interprétation des résultats obtenus au cours des années dans ce secteur de recherche (discussion sur le sujet dans Lareau, 1996). L'étude des différents intermédiaires accumulés *in vivo* démontre que ces ARN seraient répliqués par un mécanisme en cercle roulant asymmétrique (fig. 1) puisqu'aucune forme monomérique (-) circulaire ou linéaire n'a été détectée dans les extraits d'acides nucléiques de plants infectés. Les brins multimériques (-) serviraient donc directement de matrice à l'ARN pol II pour la synthèse d'ARN multimériques (+). Ces derniers doivent ensuite être coupés, et les monomères ainsi produits circularisés, afin de compléter le cycle de réplication. Il existe malheureusement très peu de résultats concluant afin d'éclaircir les mécanismes de coupure et de ligation chez les viroïdes du groupe B (revue par Flores et al., 1997). En effet, aucun consensus ne semble ressortir suite aux nombreuses années de recherche, autant au niveau de la localisation précise du site de coupure/ligation que du système enzymatique requis (Hammond et al., 1989). Par exemple, on a toujours pensé que ces viroïdes utilisaient la machinerie enzymatique de l'hôte pour permettre la coupure et la ligation des intermédiaires de réplication. Cependant, une controverse a récemment vue le jour suite aux travaux du groupe de Symons (Liu et Symons, 1998). En effet, ces auteurs ont reconstitué un essai *in vitro* permettant l'autocoupure (en absence de protéines) des brins multimériques chez CCCVd, un membre du groupe B. À l'instar des introns autocatalytiques des groupes I et II, la stabilisation d'une structure catalytique *in vitro* nécessite l'emploi de conditions non physiologiques (i.e. concentration en sels élévée). Ceci suggère que l'ARN interagirait avec des protéines cellulaires pour stabiliser la structure active de ce nouveau motif catalytique *in vivo*. En plus des résultats expérimentaux, les

deux modèles pour expliquer la réaction de coupure (autocatalytique ou enzymatique) sont appuyés par des propositions de structures secondaires conservées chez les membres du groupe B. À l'heure actuelle il est donc difficile de spéculer sur la validité de chacun des deux mécanismes proposés pour la maturation des brins multimériques.

1.3.2 Viroïdes du groupe A

Les viroïdes du groupe A ont été très peu étudiés et on ne connaissait que deux espèces appartenant à cette famille, soit ASBVd, et PLMVd. Un troisième membre de cette famille, CChMVd, a récemment été identifié (Navarro et Flores, 1997). Ces ARN ne possèdent aucune séquence, structure, ou organisation génomique similaires à ce que l'on retrouve chez les viroïdes du groupe B (pas de domaines CCR, P etc.). Les viroïdes du groupe A sont caractérisés par la présence de motifs catalytiques en tête de marteaux (*hammerhead*) qui seraient responsables de l'autocoupure des brins multimériques produits lors de la réplication (décrits dans la section 2). Il est à noter que la terminologie anglaise *hammerhead* sera utilisée dans le texte, puisqu'elle est universellement répandue dans la littérature scientifique. Mis à part l'existence de ces motifs catalytiques, les viroïdes du groupe A ne possèdent aucun autre élément de séquence ou de structure en commun. Leur regroupement repose *a priori* sur la présence de motifs catalytiques et l'absence des cinq domaines génomiques qui caractérisent les viroïdes du groupe B. L'absence de similarité au niveau des séquences rend à toute fin pratique impossible un alignement convainquant nécessaire à la réalisation des études phylogénétiques. Le regroupement de ces ARN à l'intérieur d'un même groupe (Hernandez et Flores, 1992) pourrait donc être principalement causé par l'absence d'homologies entre ces ARN et les autres viroïdes. En effet, certains types d'analyses tendent à regrouper les espèces très divergentes à l'intérieur d'un même groupe. Ce phénomène est appelé l'attraction des longues branches en phylogénie.

Bien que CarSV ait été initialement classé à l'intérieur du groupe A, certaines particularités laisse planner un doute quant à sa nature même de viroïde (Daros et Flores, 1995). En effet, les viroïdes sont des ARN infectieux qui peuvent se propager à des plants non infectés. Cependant, CarSV ne pouvait être transmis à des oeillets (hôte) sains par l'inoculation de l'ARN circulaire extrait et purifié de plants infectés. Bien qu'aucun virus ne pouvait être détecté, CarSV était considéré comme un ARN satellite dépendant d'une infection virale encore inconnue. La sensibilité du PCR a ensuite été mise à profit et à permis de montrer que CarSV existait en fait sous forme d'ADN double brin extrachromosomal (non intégré dans le génome). CarSV est retrouvé sous forme de concatémères de taille variable et associé à des séquences du pararétrovirus CERV infectant le même hôte (CERV: *carnation etched ring caulimovirus*). Lors de la réplication du virus, il y aurait eu recombinaison entre le génome viral et la forme "primitive" du viroïde CarSV (recombinaison par saut de la polymérase (*template switching*)). La structure multimérique de CarSV sous forme d'ADN double brin semble se répliquer de facon autonome et la présence de CERV ne semble pas nécessaire à sa multiplication. CarSV a donc été désigné comme étant un pararétroviroïde et les adaptations ayant été engendrées par son passage sous forme d'ADN (e.g. mécanisme de réplication, transmission) ont probablement mené à une importante divergence de cette espèce par rapport aux autres viroïdes du groupe A. L'ARN est donc non-infectieux et l'ADN extrachromosomal est transmis verticalement aux autres générations lors de la reproduction du plant hôte.

Comme pour l'ensemble de la biologie des viroïdes du groupe A, leurs mécanismes de réplication s'avère être particulièrement nébuleux. Les études menées chez un seul membre de cette famille (ASBVd) suggèrent qu'il soit répliqué par un mécanisme en cercle roulant symmétrique impliquant l'autocoupure des brins multimériques (+) et (-) par l'intermédiaire du motif catalytique *hammerhead* (section 2). Sa caractérisation chez ASBVd a d'ailleurs mené à la découverte de ce type d'ARN catalytique en 1986 (Hutchins et al.,

1986; Forster et Symons, 1987). Les études visant à identifier l'ARN polymérase impliquée dans la réplication des viroïdes du groupe A sont peu nombreuses et peu concluantes. Des expériences de réplication *in vitro* menées sur ASBVd démontrent que cet ARN serait répliqué par une ARN polymérase insensible à l'α-amanitine, éliminant les ARN polymérases II et III à titre de candidats potentiels (Marcos et Flores, 1992). De plus, des essais *in vitro* menés sur PLMVd avec l'ARN pol II de germe de blé corroborent les résultats mettant en doute l'implication de cette dernière dans la réplication des viroïdes du groupe A (revue dans Lareau, 1996). D'autres systèmes transcriptionnels sont donc à considérer (e.g. ARN pol I ou RdRp). La localisation cellulaire de ASBVd supporte l'hypothèse voulant qu'aucune des ARN polymérases nucléaires ne soit responsable de sa réplication. En effet, des études d'hybridation *in situ* sur des feuilles d'avocatier infectés ont démontré qu'il est principalement accumulé au niveau des chloroplastes (Lima et al., 1994). Sa localisation nous permet donc d'entrevoir l'utilisation de l'un ou l'autre des systèmes enzymatiques responsable de la transcription des gènes chloroplastiques. La localisation différente des viroïdes des groupes A et B amènerait donc ces ARN à interagir avec des composantes cellulaires complètement différentes et pourrait expliquer l'absence de conservation de séquences et de structures entre ces deux familles. De plus, l'acclimatation de la forme ancestrale des viroïdes à des organelles différentes pourrait avoir jouer un rôle fondamental dans la spéciation de ces ARN pour ainsi former les deux groupes tels qu'on les connait aujourd'hui. Il est cependant primordial de savoir si l'accumulation de ASBVd au niveau chloroplastique constitue une caractéristique fondamentale des viroïdes du groupe A et non une particularité liée à cette espèce (voir section 4.1).

1.3.3. Autres ARN apparentés aux viroïdes

1.3.3.1 Rétrotransposons

Une répétition en tandem que l'on croit être un rétrotransposon est retrouvée dans l'ADN répétitif (satellite 2) de certains amphibiens (Zhang et Epstein, 1996). La taille de l'unité monomérique est d'environ 300 pb. Ces régions sont transcrites par l'ARN pol II et la présence de monomères *in vivo* suggère une maturation des transcrits dimériques (Coats et al., 1994). Cette coupure est médiée par les motifs catalytiques *hammerhead* que l'on retrouve à l'intérieur de ces ARN (Epstein et Gall, 1987) (décrits dans la section 2). La caractérisation des monomères accumulés *in vivo* suggère cependant que, dans certains tissus, leur présence serait le résultat d'une initiation interne de la transcription (Epstein et Coats, 1991). En raison de sa petite taille et de la présence de motifs *hammerhead*, cet ARN (appelé rNS2T) est considéré comme étant apparenté aux viroïdes.

On a récemment caractérisé un nouvel élément d'ADN répétitif (Smaα) chez les Schistosomes qui contient un motif catalytique *hammerhead* (Ferbeyre et al., 1998). Les Schistosomes sont une famille de parasites qui infectent plusieurs espèces animales. Par exemple, trois groupes de schistosomes infectent plus de deux cent millions d'êtres humains. Les éléments Smaα sont transcrits *in vivo* par l'ARN pol III et retrouvés principalement sous forme monomérique (environ 335nt). Les motifs catalytiques sont responsables de l'autocoupure des transcrits multimériques. Les éléments répétitifs Smaα seraient également des rétrotransposons et la maturation des transcrits joueraient un rôle essentiel dans la régulation de leur transposition (Ferbeyre et al., 1998). Ce type de rétrotransposons possédant des *hammerhead* semble répandu dans d'autres espèces (e.g. criquet) et on ignore actuellement leur degré de parenté évolutive (Bourdeau et Cedergren, communication personnelle).

De par leur taille et leur nature, rNS2T et Smaα semblent être des rétrotransposons homologues. De plus, la structure et l'organisation de leurs *hammerhead* comporte d'importantes similitudes. Ces éléments ne partagent cependant aucune homologie au niveau de leur séquence, ils ne sont pas exprimés par le même système transcriptionnel et les organismes "hôtes" ne sont pas reliés au niveau évolutif. Ces arguments poussent certains auteurs à suggérer que rNS2T et Smaα ne seraient pas issus d'un ancêtre commun (Ferbeyre et al., 1998). Tout comme les viroïdes du groupe A et certains ARN satellites, ces ARN de petite taille (~300nt) utilisent des motifs *hammerhead* pour assurer la coupure de leurs intermédiaires multimériques. Ces similitudes laissent croire qu'il existe un lien évolutif entre viroïdes, ARN satellites et rétrotransposons, ce qui ravive les discussions sur leur origine et leur évolution. La question est de savoir si le *hammerhead* présent dans le génome de ces ARN est issus d'une origine évolutive commune (divergence) ou bien si il est apparu de façon spontanée dans plusieurs lignées différentes (convergence).

1.3.3.2 Virus delta de l'hépatite B humaine

Le virus delta de l'hépatite humaine (VDH) s'avère être le dernier ARN que l'on associe aux viroïdes. Le VDH est un virus satellite puisqu'il dépend de la présence d'un virus associé (virus de l'hépatite B, VHB) pour sa réplication. Son génome, dont la taille est d'environ 1700 nt, est un ARN simple brin circulaire possédant un seul cadre de lecture ouvert qui code pour l'antigène delta (Mercure et al., 1997). Cet ARN est répliqué par un mécanisme en cercle roulant, nécessitant la coupure des brins multimériques (+) et (-) (fig. 1). Cette activité est assurée par la présence d'un motif catalytique différent de ceux retrouvés chez les viroïdes et ARN satellites de virus de plantes (voir section 2). Le mode de réplication, l'existence d'un motif catalytique ainsi que la circularité de son génome sont autant d'éléments que le VDH partage avec certains viroïdes et ARN satellites. Le génome du VDH

est divisé en deux régions, l'une contenant les motifs catalytiques, et l'autre la région codant pour l'antigène delta. Récemment une protéine humaine homologue à l'antigène delta, la protéine DIPA, a été identifiée. On croit qu'un évènement de recombinaison aurait permis à un ARN de type viroïde de capturer le transcrit de DIPA pour ainsi former VDH sous sa forme actuelle (Brazas et Ganem, 1996).

2. LES MOTIFS CATALYTIQUES CHEZ LES VIROIDES ET ARN APPARENTÉS

Les viroïdes du groupe A, certains ARN satellites de virus de plantes, le VDH et les rétrotransposons rNS2T et Smaα possèdent tous des motifs catalytiques. Ce sont de courtes régions qui catalysent l'autocoupure de l'ARN (en absence de protéines), de façon spécifique. Ils sont nécessaires à la coupure des brins multimériques produits lors de la transcription (pour rNS2t et Smaα) ou de la réplication en cercle roulant (fig. 1). On en retrouve trois types différents; soit le *hammerhead* (fig. 3), le motif en épingle à cheveux (*hairpin*, pour une revue consulter Berzal-Herranz et al., 1993) et l'ARN catalytique *delta* (pour une revue consulter Mercure et al., 1997). Les deux sections qui suivent décrivent le motif *hammerhead* retrouvé à l'intérieur du génome de ces ARN.

2.1 Séquences du motif *hammerhead* (revue par Vaish et al., 1998)

Le hammerhead est illustré à la figure 3. Il est composé de trois courtes séquences conservées qui forment une structure tertiaire permettant une coupure spécifique du squelette phosphodiester suite à la formation des trois courtes hélices (tiges I, II et III). La réaction d'autocoupure dépend donc de l'assemblage d'une structure active qui requiert la présence

d'ions bivalents (e.g. Mg^{2+}). On attribuait autrefois un rôle catalytique à l'ion bivalent. Il semble aujourd'hui que sa fonction serait exclusivement structurale (Walter et al., 1998). La séquence consensus du triplet présent au site de coupure est NUH, ou N est A, C, G, ou U et H représente A, C ou U, jamais G. On connaît bien l'influence de la séquence du triplet du site de coupure (NUH) sur les différents paramètres cinétiques. Par exemple le triplet GUC confère au *hammerhead* un kcat plus élévé que le triplet GUA. De plus, on a récemment étendue la portée du triplet qui devient maintenant NHH (Kore et al., 1998). Si le $U_{16.1}$ du triplet est remplacé par un A, on doit muter le $A_{15.1}$ en U pour rétablir les ponts hydrogène entre les deux bases. De même, si l'on veut introduire un C au centre du triplet, le $A_{15.1}$ doit être cette fois remplacé par un I (inosine). Le domaine catalytique CUGANGA est quant à lui invariable, bien que l'on ait rapporté un mutant actif à l'intérieur du *hammerhead* chez un viroïde (U_4 en C, Ambros et Flores, 1998). La présence d'une pyrimidine à la position 7 est connue pour augmenter la constante catalytique. Finalement, le bloc GAAAN peut être modifié de façon à permettre les interactions nécessaires avec le triplet NHH (décrits précédemment).

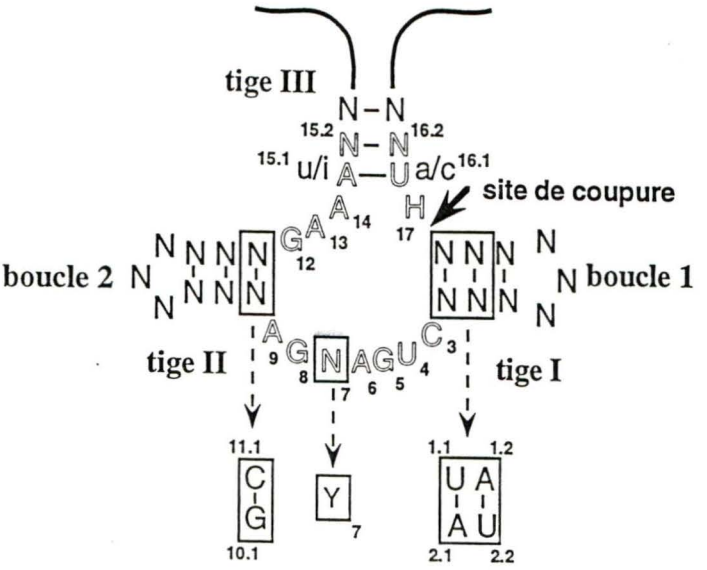


Figure 3. Séquences et structures du *hammerhead*. Représentation de la structure secondaire du *hammerhead*. Les lettres doubles montrent les séquences des trois domaines conservés. Les mutations possibles des positions 15.1 et 16.1 des domaines NUH et GAAAN sont indiquées en minuscule. Les encadrés montrent les positions ainsi que l'identité des nucléotides qui augmentent la vitesse de la réaction catalysée par le *hammerhead*. La numérotation utilisée suis la proposition de Hertel et collaborateurs (Hertel et al., 1992).

Ayant très tôt reconnu et caractérisé l'importance des régions conservées, on a toujours pensé que la réaction catalysée par le *hammerhead in vitro* était peu influencée par la séquence et la longueur de ses tiges. Cependant, les groupes de Uhlenbeck et Eckstein ont mis en évidence des éléments à l'intérieur des tiges I et II qui peuvent jouer un rôle significatif sur la vitesse de la réaction (kcat). Au niveau de l'identité des nucléotides, on a réalisé l'importance de l'appariement $G_{10.1}$-$C_{11.1}$ comme première paire de base de la tige II (Tuschl et Eckstein, 1993). La substitution du reste de la tige par une courte boucle de quatre U, par exemple, n'a que peu d'effet sur la vitesse de coupure. Quant à la longueur des tiges, on a observé un phénomène d'inhibition lorsque les tiges I et II dépassent une taille critique (5 pb pour la tige I et 4 pb pour la tige II). La combinaison d'une longue tige I et d'une courte tige II (ou *vice versa*) ne provoque aucune inhibition et le ribozyme peut bénéficier de la combinaison des appariements $U_{1.1}$-$A_{2.1}$ et $A_{1.2}$-$U_{2.2}$ au début de la tige I qui augmente de dix fois la vitesse de la réaction (Clouet-d'Orval et Uhlenbeck, 1997).

2.2 Organisation des motifs *hammerhead*

La réaction d'autocoupure médiée par le *hammerhead* nécessite l'adoption d'une structure catalytique active. Elle peut ainsi être inhibée lorsque le motif catalytique est dissimulé à l'intérieur de longs ARN hautement structurés, comme le génome des viroïdes. En effet des structures alternatives très stables peuvent interférer dans le repliement du motif catalytique. Cet équilibre structural peut affecter de façon importante l'efficacité du *hammerhead*. La formation transitoire du *hammerhead* durant la réplication pourrait être un mécanisme permettant la coupure de ces ARN.

Chez les ARN "naturels" (e.g. viroïdes), les trois blocs conservés formant le *hammerhead* sont pratiquement toujours consécutifs. En effet, l'ARN extracatalytique (la partie du génome qui ne forme pas le motif catalytique) est situé dans la région formant soit la tige I ou III du *hammerhead* (fig. 4, *hammerhead* A ou B, respectivement). Cependant chez certaines espèces comme ASBVd, l'ARN extracatalytique se retrouve à la fois dans la région des tiges I et II, séparant ainsi le domaine catalytique (CUGANGA) du reste du *hammerhead* (fig. 4, *hammerhead* C). De plus, chez les ARN ayant des *hammerhead* de configuration A ou C (fig. 4), on observe une tige III qui est considérée comme peu stable en raison de sa longueur réduite (2 pb). Certains auteurs ont suggéré que cette contrainte rend peu probable la formation de la structure catalytique conventionnelle et ont proposé un mécanisme alternatif en double *hammerhead* (voir fig. 4) (Forster et al., 1988). Il s'agit d'une interaction en *trans* entre deux *hammerhead* qui s'associent de facon à allonger, et donc stabiliser, la tige III. Les deux motifs catalytiques sont formés simultanément (d'ou le nom double *hammerhead*) et chaque ARN est mutuellement coupé par l'autre ARN. Cette caractéristique différencie ce modèle d'une simple réaction en *trans* ou le *hammerhead* de l'ARN #1 reconnaît, lie et coupe le site de coupure du *hammerhead* de l'ARN #2 sans qu'il n'y ait adoption concomittante des deux structures catalytiques. Cependant, il semble que les *hammerhead* à tige III très courte soient également capable d'autocoupure en utilisant un mécanisme conventionnel en simple *hammerhead* (réaction en *cis*) (Epstein et Pabon-Pena, 1991). Par exemple une tige III de 2 pb avec une boucle de 4 nt ou une tige de 3 pb et une boucle de 3 nt donne une stabilité suffisante pour permettre la formation d'une telle structure (Sheldon et Symons, 1989). La coupure en *cis* ou celle en *trans* pourraient dépendre du contexte structural de l'ARN à l'intérieur duquel se retouve le *hammerhead*.

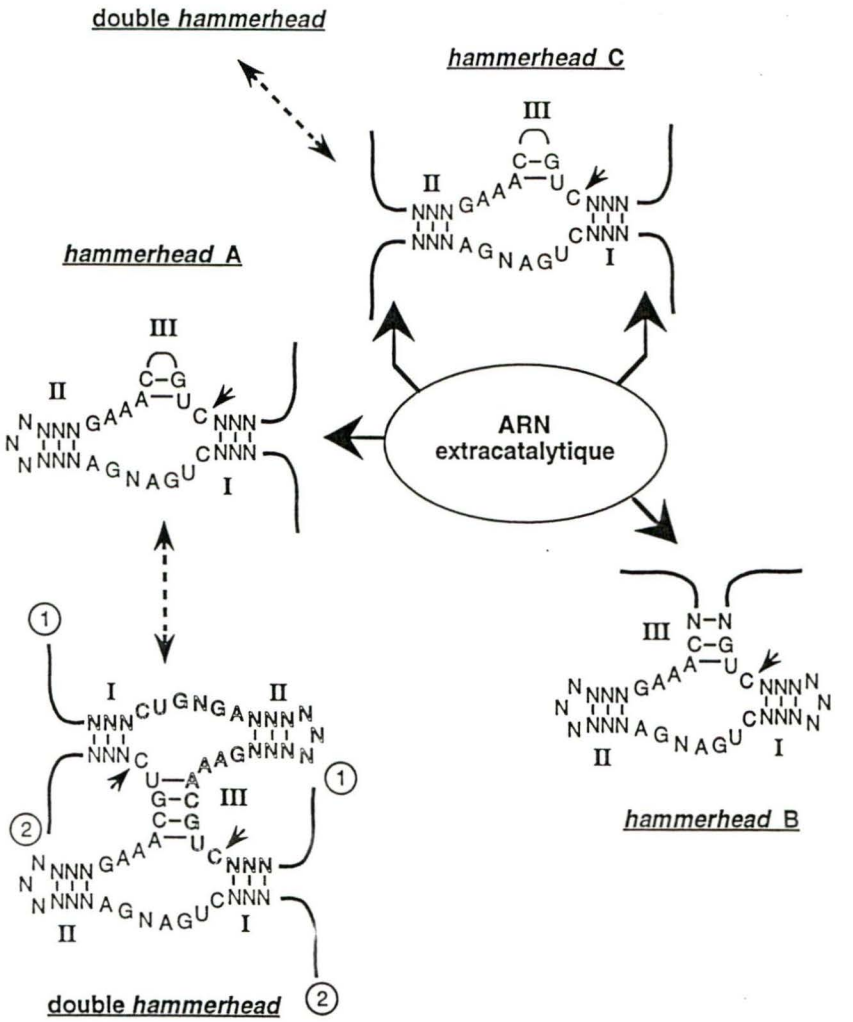


Figure 4. Organisation des *hammerhead*. La localisation de l'ARN extracatalytique dans la région de la tige I (hammerhead A), III (hammerhead B) ou I et II (hammerhead C) du *hammerhead* est schématisée. L'instabilité de la structure des *hammerhead* A et C pourrait favoriser la formation d'une structure en double *hammerhead* impliquant une interaction en *trans* entre deux ARN.

La découverte du *hammerhead* dans les années 80 a rapidement mené à l'élaboration des ribozymes, enzymes à base d'ARN. Si on linéarise le *hammerhead* A de la figure 4 dans la boucle de la tige I, on obtient un ribozyme et son substrat. Le substrat est constitué du site de coupure (NUH) et des deux régions adjacentes qui permettaient la formation des tiges I et III. Le ribozyme est quant à lui formé du domaine catalytique (CUGANGA), du bloc GAAAN et des régions impliquées dans la formation des tiges I et III, qui deviennent les sites de reconnaissance et de liaison du substrat. En modifiant leur séquence, on peut lier [illegible] et catalyser la coupure d'ARN cibles (substrats) contenant un site de coupure NUH. Suite à la coupure, les produits sont relâchés et une deuxième molécule d'ARN peut se fixer au ribozyme. On obtient ainsi un *turnover*, caractéristique fondamentale d'une enzyme.

3. L'ÉVOLUTION DES VIROIDES ET DES ARN APPARENTÉS

Au début des années 90 le groupe de Diener réalisait une étude phylogénétique incluant 15 espèces de viroïdes (14 du groupe B et ASBVd du groupe A), 6 ARN satellites et le domaine incluant les motifs catalytiques du VDH (Elena et al., 1991). Leurs résultats montrent que les viroïdes et les ARN apparentés (VDH et ARN satellites) forment deux groupes monophylétiques distincts. L'arbre phylogénétique permet donc de séparer les ARN à réplication autonome qui sont incapables de s'autocouper (viroïdes du groupe B), des ARN à réplication non-autonome et dotés de motifs catalytiques (ARN satellites). ASBVd qui possède des caractéristiques fontionnelles des deux types d'ARN est situé entre les deux groupes sur l'arbre phylogénétique. Il est donc considéré comme un lien évolutif entre les ARN satellites et les autres viroïdes. Cette hypothèse a été ensuite renforcée par la découverte de PLMVd, un viroïde ayant des motifs *hammerhead* et qui occupe une position semblable dans l'arbre phylogénétique (Hernandez et Flores, 1992). ASBVd et PLMVd

appartiennent à une classe de viroïdes très différents et forment le groupe A. Les viroïdes du groupe B apparaissent quant à eux subdivisés en deux sous-groupe (B1 et B2) en fonction de l'identité de la séquence de leur domaine CCR. On constate finalement une division du sous-groupe B1 qui coïncide avec la taille du génome, les viroïdes de plus de 300 nt et ceux de moins de 300 nt.

Les études sur l'évolution nous amènent bien sûr à s'interroger sur l'origine des viroïdes et des ARN apparentés. Cette question a engendré de nombreuses hypothèses au cours des années. Certains suggèrent que les viroïdes sont des reliques d'ARN primitifs ayant précédé l'apparition de la vie sur terre (revue par Diener, 1989). L'hypothèse d'un monde à ARN suggère en effet que l'ARN aurait été le premier système génétique capable de s'autorépliquer et d'évoluer (revue par Orgel, 1994 et Joyce, 1989). L'ARN aurait précédé l'ADN dont le potentiel catalytique est plus limité (voir aussi Bussière et Perreault, 1995; Annexe 1.2). Ce concept s'appuie également sur les études de la chimie prébiotique qui laissent croire que les acides ribonucléiques aient été les premières macromolécules à avoir été synthétisées (revue par Orgel, 1994). Les viroïdes ayant une information génétique et des capacités catalytiques (phénotype) seraient des exemples du type de molécule ayant peuplé la terre à cette époque. Leur structure circulaire faciliterait la réplication qui peut être initié à n'importe quel endroit sans qu'il n'y ait perte de l'information génétique. Celle-ci serait aujourd'hui assurée par un système enzymatique protéique. D'autres, ayant constater certaines homologies de séquences et de structure entre les viroïdes et les introns du groupe I, ont proposé que les viroïdes seraient des introns ayant acquis une réplication autonome (revue par Diener, 1989). Finalement, l'existence des motifs *hammerhead* chez certains viroïdes suggère qu'ils auraient évolué à partir d'ARN satellites de virus de plantes ou bien de rétrotransposons apparentés aux éléments rNS2T et Smaα. La présence de motifs catalytiques demeure la pierre angulaire de ces hypothèses et leur caractérisation dans divers

systèmes biologiques s'avère essentielle à une meilleur compréhension des relations évolutives entre ces ARN.

4. OBJECTIFS DE LA RECHERCHE

4.1 La biologie des viroïdes du groupe A

4.1.1 Contexte

Depuis la découverte des viroïdes dans les années 70, les études visant à mieux comprendre leur biologie ont principalement porté sur les espèces du groupe B. En 1994, on ne connaissait que deux viroïdes appartenant au groupe A, soit ASBVd et PLMVd. Ces viroïdes possèdent des caractéristiques bien différentes de celles retrouvées chez ceux du groupe B. Cependant, tel qu'illustré à la figure 5, les différents aspects de leur biologie moléculaire demeurent méconnus. Mis à part la présence des motifs catalytiques *hammerhead*, aucun élément de séquence ou de structure conservé n'a été identifié chez les viroïdes du groupe A. Cette situation contraste avec celle des autres viroïdes chez lesquels on a identifié de nombreuses structures conservées auxquelles on attribue parfois certaines fonctions biologiques (voir section 1.3.1 et fig. 2) (Flores et al., 1997). Les quelques études menées sur ASBVd montre que cet ARN est situé au niveau des chloroplastes des cellules infectées (Lima et al., 1994) où il serait répliqué par un mécanisme en cercle roulant symmétrique (Bruening et al., 1982). On connaît cependant peu de chose sur l'identité de l'ARN polymérase responsable de la réplication des viroïdes du groupe A. Chez ASBVd, des études d'inhibition avec l'α-amanitine suggèrent l'implication d'un système enzymatique différent de celui utilisé par les viroïdes du groupe B. Sa localisation cellulaire

laisse entrevoir un rôle potentiel des sytèmes transcriptionnels chloroplastiques. Finalement, des expériences d'autocoupure *in vitro* suggèrent que les brins multimériques (+) et (-) produits lors de la réplication seraient coupés suite à la formation d'une structure en double *hammerhead* (fig. 4) (Forster et al., 1988; Davies et al., 1991). Cette interaction en *trans* entre deux motifs catalytiques est moins favorisée que la formation d'une structure intramoléculaire en simple *hammerhead*, ce qui réduirait l'efficacité de la réaction. Ces résultats corrèlent avec le patron d'accumulation de ASBVd *in vivo*. En effet, à l'intérieur des plants infectés, les intermédiaires de réplication correspondent à une série d'ARN dont la taille varie de une à huit répétitions de l'unité monomérique (Bruening et al., 1982). La présence de concatémères résultant d'une autocoupure incomplète des brins multimériques serait le fruit du mécanisme utilisé pour la formation des structures catalytiques. Cet exemple illustre bien la complémentarité des études *in vitro* et *in vivo* pour la compréhension des phénomènes biologiques.

4.1.2 PLMVd: un modèle

ASBVd fait en général figure d'exception puisqu'il est le seul de son groupe à avoir été substantiellement étudié. Il est par conséquent risqué de tirer des conclusions pertinentes à la compréhension du cycle infectieux des viroïdes du groupe A en se basant sur les observations faites sur un seul de ses membres. Des études portant sur d'autres viroïdes apparentés à ASBVd sont donc primordiales. La caractérisation de PLMVd (Hernandez et Flores, 1992) s'avère être une occasion de faire avancer les connaissances dans ce domaine de recherche. PLMVd est le viroïde responsable de la mosaïque latente du pêcher (Flores et al., 1990). La maladie induite par PLMVd chez les arbres infectés est généralement latente la première année. Durant la deuxième année apparaissent les premiers symptômes: délai du développement foliaire et floral, taches sur les feuilles, puis détérioration des fruits. Après la cinquième année, les arbres vieillissent rapidement et deviennent plus sensibles au gel et aux

maladies (Flores et al., 1990). Ce viroïde, dont la taille est d'environ 338 nt, possède des motifs autocatalytiques *hammerhead* permettant l'autocoupure des brins multimériques (+) et (-) produits lors de la réplication en cercle roulant (Hernandez et Flores, 1992). PLMVd possède des motifs *hammerhead* qui sont organisés de façon très différentes de ceux retrouvés chez ASBVd (fig. 4, comparer hammerhead B et C). Les séquences extracatalytiques sont exclusivement situées dans la région de la tige III, ce qui suggère une réaction d'autocoupure en simple *hammerhead* (pour les deux polarités). Ces quelques notions résument l'essentiel des connaissance acquises sur PLMVd. Les objectifs de nos recherches visant à une meilleure connaissance de la structure et des mécanismes de réplication des viroïdes du groupe A sont présentés ci-dessous. Ils sont également illustrés à la figure 5. PLMVd sera utilisé comme modèle d'étude.

4.1.3 Objectif général

- Caractériser différents aspects de la **biologie moléculaire de PLMVd** pour mieux comprendre les différentes étapes de son cycle de réplication et, par conséquent, celui des **autres viroïdes du groupe A**.

4.1.4 Objectifs spécifiques

- Élucider les **mécanismes d'autocoupure des *hammerheads*** *in vitro* chez PLMVd (chapitre 1).

- Identifier les différentes formes du viroïde retrouvées chez les plants infectés. Nous pourrons ainsi connaître la **nature** et les **caractéristiques** du **cercle roulant** utilisé par PLMVd (chapitre 2).

- Étudier la **localisation cellulaire** de PLMVd à l'intérieur des cellules infectées (chapitre 2).

- Caractériser les **éléments structuraux** du génome de PLMVd en solution (chapitre 3).

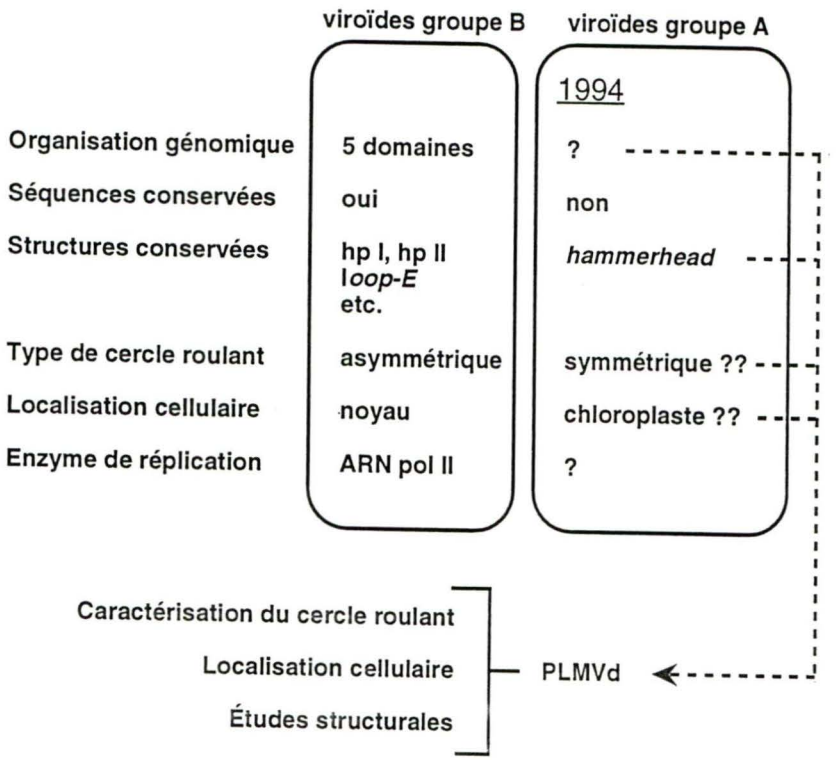


Figure 5. Comparaison des principales caractéristiques des viroïdes des groupes A et B. La figure schématise les connaissances antérieures à l'année 1994 et identifie les principales questions adressées par les travaux présentés dans ce travail. Les ? montrent des caractéristiques inconnues (?) ou qui ne sont pas nécessairement représentatives du groupe (??).

4.2 Évolution des viroïdes et des ARN apparentés

4.2.1 Contexte

Les premières analyses phylogénétiques montrent que les viroïdes et les ARN satellites forment deux groupes monophylétiques distincts (Elena et al., 1991). ASBVd, le seul viroïde à posséder des motifs autocatalytiques *hammerhead*, est considéré comme un lien évolutif entre ces ARN. Cette hypothèse est renforcée en 1992 par la caractérisation d'un deuxième viroïde ayant des *hammerhead* (PLMVd). PLMVd et ASBVd forment le groupe A dont les membres possèdent des caractéristiques à la fois des viroïdes et des ARN satellites. Les viroïdes du groupe B apparaissent quant à eux divisés en deux sous-groupes, B1 et B2. Le séquençage de CbVd1, en 1990, suggère l'existence d'un troisième sous-goupe dont on ignore la position phylogénétique (Spieker et al., 1990). Dans la dernière décennie de nouveaux viroïdes ainsi que de nombreux variants des espèces déjà connues ont été séquencés. On a, entre autres, identifié un viroïde, deux ARN satellites et plusieurs nouveaux transcrits d'ADN répétitifs qui possèdent des motifs catalytiques *hammerhead.* En l'absence de règles sur l'identification des nouveaux ARN, une nomenclature très complexe voit le jour suite à cette recrudescence du nombre de séquences connues.

4.2.2 Objectif général

- Étudier l'**évolution des viroïdes et des ARN apparentés** (chapitres 1 et 2).

4.2.3 Objectifs spécifiques

- **Répertorier toutes les séquences des viroïdes et ARN apparentés** disponibles dans les banques de données ou dans les articles scientifiques. Mettre sur pied et proposer une nouvelle **nomenclature pour identifier ces ARN**. Créer un **site web** ou cette information sera disponible pour la communauté scientifique qui oeuvre dans ce secteur de recherche (chapitre 4).

- Tenter d'obtenir une reconstruction phylogénétique ayant une **meilleure résolution des relations évolutives des espèces proches**. Profiter des nombreuses nouvelles séquences connues chez les viroïdes du groupe B pour améliorer la topologie de cette région de l'arbre phylogénétique (chapitre 4).

- Réévaluer l'**existence du groupe A** et sa position phylogénétique par rapport aux autres viroïdes et aux ARN satellites. Étudier également les **relations évolutives entre viroïdes, ARN satellites et autres ARN apparentés**. La caractérisation des nouveaux ARN ayant des motifs *hammerhead* sera mise à profit (chapitre 5).

RÉSULTATS

CHAPITRE 1:

Autocoupure des *hammerhead* chez PLMVd

1.1 ARTICLE:

The RNA of both polarities of the peach latent mosaic viroid self-cleaves *in vitro* solely by single hammerhead structures

Danièle Beaudry, **Frédéric Bussière**, François Lareau,

Carl Lessard et Jean-Pierre Perreault

Département de Biochimie

Université de Sherbrooke

Sherbrooke, Québec, J1H 5N4, Canada

Article publié dans: *Nucleic Acids Research*, Vol. 23, 745-752, 1995.

ABSTRACT

Hammerhead self-cleavage of dimeric, monomeric, truncated and mutated transcripts derived from both polarities of the peach latent mosaic viroid (PLMVd) were characterized. In contrast to some results previously published for a very close sequence variant (see ref.;1), these RNAs exhibit a virtually identical self-cleavage during transcription and after purification. By self-cleavage of dimeric transcripts with normal and mutated hammerhead domains and by complementation experiments, we show that the cleavage reactions involve only single hammerhead structures. This observation contrasts with the case of avocado sunblotch viroid (ASBVd), the other self-cleaving viroid, whose mechanism involves mostly double hammerhead structures, whereas single hammerhead cleavage is associated with viroid-like plant satellite RNAs. The difference of stability between the native secondary structures adopted by viroids and the autocatalytic structures including the hammerhead motif governs the efficiency of the self-cleavage reaction. The transition between those conformers is the limiting step in catalysis and is related exclusively to the left arm region of PLMVd secondary structure, which includes the hammerhead sequences. Most of the mutations between the variant we used and the sequence variant previously published are located in this left arm region, which may explain to a great extent the differences in their cleavage efficiency. No interactions with long-ranged sequences contributing to the autocatalytic tertiary structure were revealed in these experiments.

INTRODUCTION

Viroids are small single stranded circular RNAs that infect higher plants, causing diseases in crop species and important economic losses in agriculture (2,3). It was proposed that viroids replicate in a DNA-independent manner by a rolling circle mechanism which involves synthesis of multimeric strands that have to be cleaved into monomeric fragments before being circularized in order to yield the progeny (4). Viroid-like plant satellite RNAs (also named virusoids) and two viroids (ASBVd and PLMVd) undergo specific self-cleavage of multimeric strands into monomeric strands with their autocatalytic "hammerhead" or "hairpin" structures (3). The molecular mechanism of these autocatalytic domains has been extensively studied using minimal structures (3, 5). It is not excluded that cellular factors are also involved in the processing of multimeric strands. For example, ribonuclease T1 was shown to cleave and ligate *in vitro* more than one unit-lenght potato spindle tuber viroid (PSTVd) transcripts (6), which are not known to possess any autocatalytic sequences. The self-cleavage is probably an ancestral reaction and the cleavage involving cellular factors would derive from this primitive mechanism (7).

The ability of viroids that possess autocatalytic sequences appear to depend on the capacity of the RNA to adopt a conformation different from their most stable native structure. The hammerhead self-cleavage occurs by single or double hammerhead structure, depending on the ability of the sequence to form stable stems surrounding the catalytic site, especially stem III (Fig. 1A-B; see Ref. 3,8). For example, the plus strand of ASBVd requires the association of two hammerhead structures to stabilize the catalytic core for self-cleavage to occur during and after *in vitro* transcription (8,9). In contrast, the minus strand of ASBVd also self-cleaves by double hammerhead structure during transcription but mostly by single hammerhead after gel purification (9). In addition, a transcript of satellite 2 DNA from the newt self-cleaves by alternative single and double hammerhead modes (10). Unlike viroids, satellite RNAs, whose hammerheads contain stable stem III, are likely to undergo intramolecular self-cleavage by single structure as shown for the satellite RNA of lucerne transient streak virus (vLTSV), which performs self-cleavage uniquely by single hammerhead structures (8).

The peach latent mosaic viroid (PLMVd) is the causal agent of PLM disease (12). Like the viroid-like satellite RNAs, the hammerhead structures of both polarities from PLMVd have a stable stem III (Fig. 1C-D) suggesting that self-cleavage occurs via single structures. Hernandez and Flores (1) have shown that self-cleavage occurs efficiently during transcription from a monomeric PLMVd construction, but that after purification, self-cleavage occurred at a reduced level, on both strands. Consequently they proposed that the hammerhead structure is more easily adopted during transcription than after a complete synthesis. Using a PLMVd sequence variant (Fig. 1E), we studied several features of transcript self-cleavage in order to establish clearly the mode of hammerhead mechanism and to identify the features regulating the efficiency of the reaction. Unlike several studies using minimal structures, we used a complete viroid sequence because the results obtained with the entire genome reflect more accurately the *in vivo* replication process.

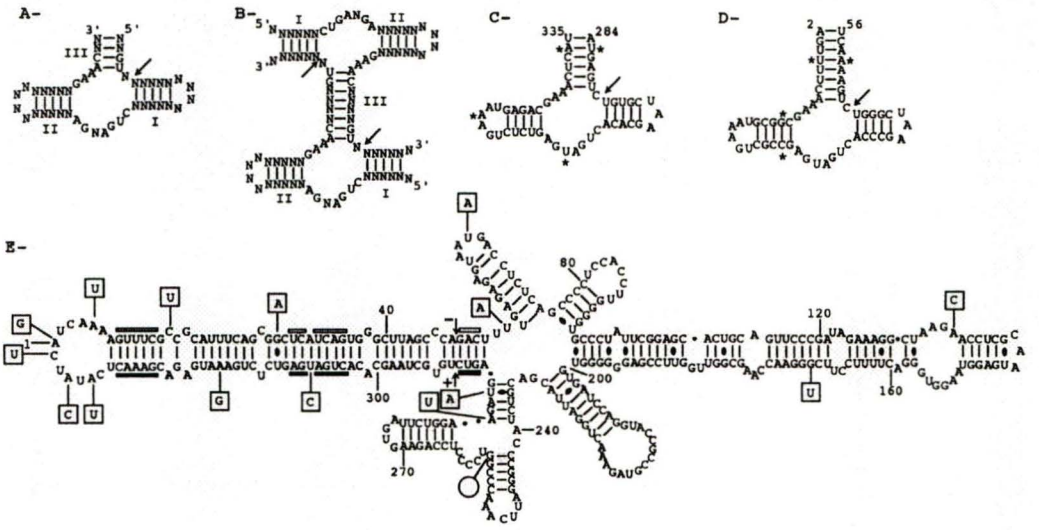


Figure 1. Proposed secondary structures. A- and B- are single and double hammerhead consensus secondary structures. Arrows indicate the cleavage sites. C- and D- are single hammerhead structures proposed for both plus and minus polarities of PLMVd. (*) indicate the position of nucleotide differences between the sequence variants characterized. E- The lowest free energy secondary structure proposed for the PLMVd sequence used in the present study. The secondary structure was determined using the 2.0 version of the MFold program based on Zuker (12). Nucleotides in the sequence variant reported by Hernandez and Flores (1) that differ from the variant used here are in boxes, and the additional G is indicated by a circle. Hammerhead consensus sequences are indicated by an open bar for minus polarity and by a closed bar for plus polarity.

MATERIALS AND METHODS

Reagents

Restriction enzymes, DNase I (RNase free), calf intestinal alkaline phosphatase, T4 DNA ligase, RNA guard, T7 sequencing kit and Sephadex G-50 were purchased from Pharmacia. TAQ DNA polymerase, T3 and T7 RNA polymerase were from Promega (Fisher Scientific). [α^{32}P]-UTP (3000 Ci/mmol) and [α-^{35}S]-dATP (1000 Ci/mmol) were from Amersham Canada.

PLMVd plasmids

pPL5 clone was a generous gift from Dr. Ricardo Flores. pPL5 clone is a recombinant plasmid derived from pBluescript II KS (+/-) (Stratagene) which contains at the Pst I site of the polylinker a monomeric Pst I insert of PLMVd initialy cloned in pSPT 18 (Fig. 2;1). The pPL5 insert is a sequence variant of that used by Hernandez and Flores (1) in their original PLMVd self-cleavage characterization. For the pPD1 clone, the PLMVd insert of pPL5 clone was digested by Pst I and gel purified on a 1% agarose gel. Dimeric fragments resulting from the ligation (1:1 ratio) of Pst I monomeric fragments were ligated into dephosphorylated pBluescript II KS (+/-) vector digested with Pst I. Both monomeric and dimeric inserts have been sequenced in either directions by the dideoxynucleotide chain-termination method using T7 sequencing kit (Pharmacia). Sequencing of several clones confirmed the presence of an additional G in position 258.

In vitro transcription and RNA purification

For preparation of complete plus strands, clones were digested with BamH I and transcribed with T3 RNA polymerase (see Fig. 2). To obtain complete minus strands, the clones were digested with EcoR I and transcribed with T7 RNA polymerase. Partial sequences of both polarities have been prepared using various restriction enzymes (see Table II). Radioactive transcripts were prepared by incubating 5 mg of linearized DNA template overnight at 37°C with 0.5 mM ATP, 0.5 mM GTP, 0.5 mM CTP, 0.01 mM UTP, 40 mCi $\alpha[^{32}P]$-UTP, 40 mM Tris-HCl pH 7.9, 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 10 mM DTT, 34 U RNA guard and 120 U T3 or T7 RNA polymerase in a volume of 100 ml. The mixtures were pre-incubated for 10 min at 37°C before adding RNA polymerase. After transcription, the mixtures were incubated for 10 min at 37°C with 30 U DNase I and 4 ml of 0.5 M EDTA pH 8.0 and then extracted with phenol:chloroform. For evaluation of self-cleavage efficiency; 0.5 vol stop buffer (0.3% bromophenol blue and xylene cyanol, 10 mM EDTA pH 7.5 and 97.5% deionized formamide) was added to transcriptional aliquots, denatured for 2 min at 65°C and analyzed on a 5% polyacrylamide gel in 100 mM Tris-borate pH 8.3, 1 mM EDTA with 7 M urea. The fraction of cleaved molecules was determined by analysis of dried gels with a PhosphorImager (Molecular Dynamics). For purification; the radioactive transcripts were precipitated before gel separation. Transcripts were detected by autoradiography and then were excised, eluted, precipitated, passed twice through Sephadex G-50 spun columns, lyophilized and conserved at -70°C. For preparation of non-radioactive transcripts; the procedure was identical that for the radioactive transcripts except that the concentration of UTP was 0.5 mM and the radioactive UTP was omitted from the pool. Products from the self-cleavage reactions were purified on 5% PAGE (7M urea), detected by UV-shadowing, extracted and their concentration was determined from their absorption in aqueous solution at 260 nm.

Synthesis of deletion-mutants

Truncated transcripts were synthetized from PCR-prepared templates where the control hammerhead GAAAC sequence was either conserved or mutated to GAAC, following the approach described previously for ASBVd self-cleavage studies (8,9). For the plus strand, both pPL5 and pPD1 plasmids were linearized with BamH I, and the viroid sequence was PCR-amplified with a sense oligonucleotide of 17 nucleotides corresponding to the T3 promoter, and an antisense oligonucleotide of 23 (5'GATATGA**GTTC**GTCTCATTTCAG3', deletion) or 24 (5'GATATGA**GTTTC**GTCTCATTTCAG3', control) nucleotides corresponding to positions 315-338 of PLMVd (see Fig. 1). PCR mixtures were 1 μg DNA, 5 pmol of each oligonucleotides, 10 mL Perkin-Elmer 10X buffer (500 mM Tris-HCl pH 8.3 and 500 mM KCl), 2.5 mM $MgCl_2$, 2.5 U TAQ DNA polymerase in a total volume of 100 mL. Amplification cycle (94°C for 2 min, 30°C for 1 min, heated by 1°C per 3 sec up to 72°C, 72°C for 1 min) was repeated only 5 times in order not to restrict the PCR product to the shortest product from the pPD1 template, which include only one hammerhead sequence. After phenol-chloroform extraction and isopronanol precipitation, PCR reactions were gel purified on a 2% agarose gel and the bands of 313 (control) and 312 (deletion) (from pPL5 and pPD1), 651 (control) and 650 (deletion) (from pPD1) nucleotides were isolated, extracted, ethanol precipitated and rinsed, dried and then *in vitro* transcribed as described above. The same approach was used for the minus strand of pPL5 and pPD1 previously linearized with EcoR I. The sense oligonucleotides of 26 (5'CATCAAAA**GTTC**GCCGCATTTCAGCG3', deletion) or 27 (5'CATCAAAA**GTTTC**GCCGCATTTCAGCG3', control) nucleotides covered the positions 338-24, while the antisense oligonucleotide corresponds to the T7 promoter (5'TAATACGACTCACTATA3'). Amplification conditions were similar to those described above and the product of 173 and 511 nucleotides for the control while 172 and 510 nucleotides for the mutant were isolated.

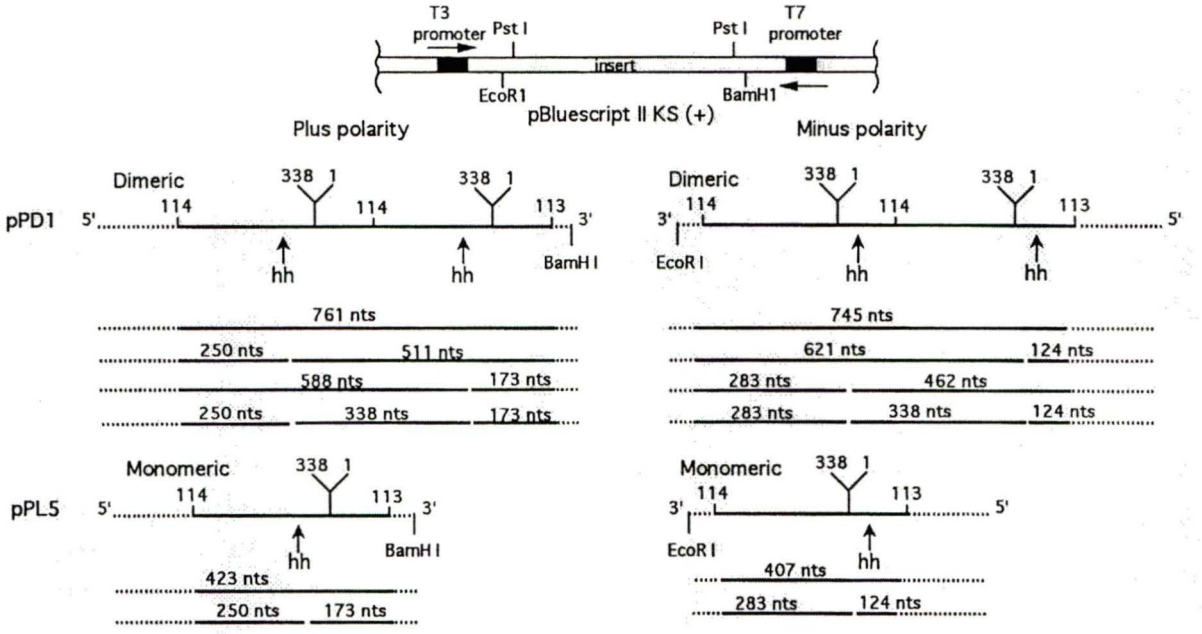


Figure 2. PLMVd plasmidic constructions and transcriptional products. Monomeric and dimeric constructions are reported in Materials and Methods. Full lines indicate sequence from PLMVd, and dashed lines sequence from the vector. Arrows denoted by hh are for hammerhead cleavage sites and the length of each fragment in nucleotides (nts) is indicated.

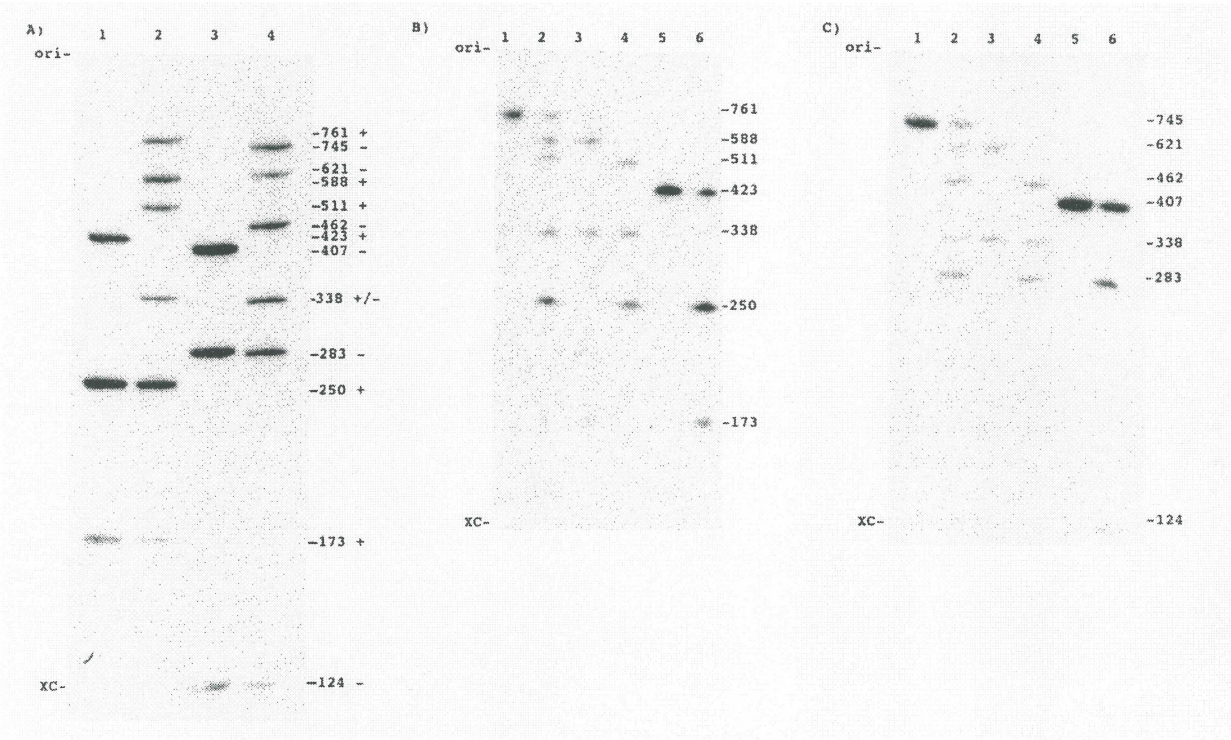


Figure 3. Evaluation of self-cleavage efficiency. A- During in vitro transcription. Lanes 1 and 2 are monomeric and dimeric plus polarity transcripts and lanes 3 and 4 monomeric and dimeric minus polarity transcripts. The length of the products is indicated in the margin with the symbol specifying the polarity. B- and C- Self-cleavage of purified plus and minus polarity transcripts, respectively. Lanes 1 and 2, full dimeric transcripts; lane 3, dimeric transcripts with a hammerhead sequence in 3'; lane 4, dimeric transcripts with a hammerhead sequence in 5'; lanes 5 and 6 the monomeric transcripts. In lanes 1 and 5, snap-cooling and MgCl2 were omitted. ori denotes origin of migration and XC xylene cyanol.

In vitro self-cleavage of purified transcripts

Prior to self-cleavage incubation, the samples (25,000-100,000 CPM) were heated in 1 mM EDTA pH 6.0 at 100°C for 1 min and snap-cooled on ice for 1 min. Self-cleavage of purified transcripts was initiated by adding the reaction buffer to a final concentration of 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ and 0.5 mM EDTA in total volume of 10 ml and then samples were incubated for 15 min at 37°C. The reactions were stopped by addition of 0.5 vol stop buffer, kept on ice, denatured for 2 min at 65°C, purified on 5% PAGE (7M urea) and the dried gels were analyzed by PhosphorImager. Variations to this protocol in some experiments are indicated further.

RESULTS

Monomeric and dimeric transcripts self-cleave similarly

pPL5 and pPD1 clones included respectively an insert of monomeric and dimeric PLMVd sequence as shown in figure 2. After linearization, these constructions allowed the synthesis of either plus or minus transcripts according to whether T3 or T7 RNA polymerase was used. During *in vitro* transcription, RNA of both plus and minus polarities self-cleaved efficiently (Fig. 3A) indicating that hammerhead structures were adopted. Once purified, complete transcripts as well as the intermediate products from dimeric transcripts, which self-cleaved at a unique site during transcription, showed autocatalytic cleavage at the same level as monomeric and dimeric transcripts (Fig. 3B-C). If Mg^{2+}, snap-cooling or both were omitted (Fig. 3B-C, lanes 1 and 5), self-cleavage could not be detected. Both monomeric and dimeric plus polarity transcripts displayed a slightly greater self-cleavage than did the minus polarity transcripts (Fig. 3); plus transcripts self-cleaved at 60-70% while minus transcripts at 50-55%. The total cleavage at one site in a dimeric transcript was obtained by addition of the percentage of product resulting from the cleavage at this site only and the one from the cleavage which occurred at both sites. Furthermore, 5' and 3' self-cleavage appear to be virtually identical within a dimeric transcript, suggesting that the

location of the hammerhead domain as well as the sequences derived from the vector do not influence the reaction efficiency.

Time courses of self-cleavage for purified transcripts were performed (Fig. 4). With the conditions we used, self-cleavage occurred almost immediately for dimeric transcripts: after 15 sec most of the substrate which could self-cleave did so (Fig. 4). Similar results were obtained at concentration ranging from 0,2 to 45 nM. Almost no additional transcripts reacted when incubations were carried from 15 to 600 sec, and even after 1 hour of incubation, the level of self-cleavage remained unchanged. Similar results with time courses for monomeric transcripts from pPL5 and intermediate transcripts with one hammerhead site from pPD1 were also observed (data not shown). In addition, the preliminary snap-cooling step was characterized for several transcripts and shown similar results. For example, the experiments using the plus polarity monomeric RNA showed that the denaturation must be at 90°C or higher to observe efficient self-cleavage. If a second cycle of snap-cooling and incubation at 37°C was performed, the self-cleavage is increased in the same proportion as during the first cycle, while after three cycles almost all the transcripts did cleave, indicating that the population is fully productive (Fig. 4C). Thus, all transcripts are reactive and the limiting step appears to be the catalytic core formation, not the cleavage step.

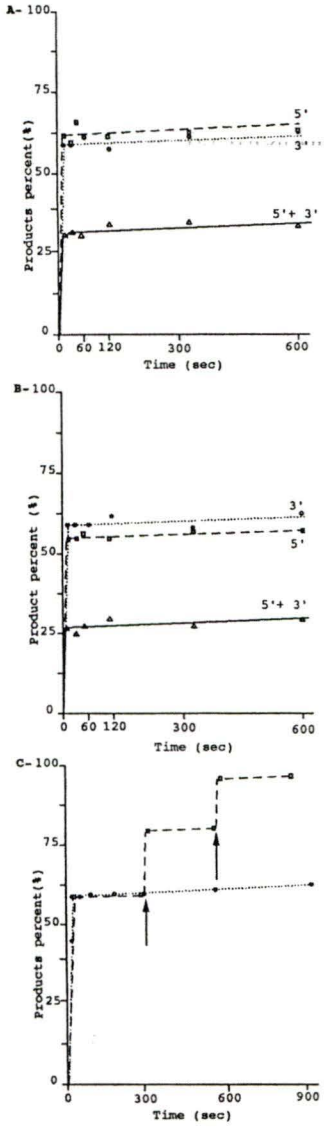


Figure 4. Self-cleavage time courses of 1 nM purified transcripts. A- and B- are time courses of complete dimeric transcripts of plus and minus polarities, respectively. Plain lines (__) are for self-cleavage occuring at both hammerhead sites, while dashed lines are for self-cleavage occuring at the 5' site (----) and the 3' site (....). C- is self-cleavage of plus polarity monomeric transcripts. Doted lines (.....) are for self-cleavage after a single snap-cooling at 90°C and dashed lines (-----) are for the case where a second and third snap-cooling (indicated by arrows) have been performed.

Only single hammerhead structure is involved in self-cleavage of PLMVd

To investigate specifically the single versus double hammerhead modes of self-cleavage, we prepared truncated monomeric RNAs of both polarities from pPL5 by PCR followed by *in vitro* transcription, with the hammerhead GAAAC sequence either conserved or mutated to GAAC, according to the strategy reported for ASBVd self-cleavage studies (8,9). For both polarities transcripts, the self-cleavage was productive during *in vitro* transcription and after purification when the GAAAC sequence was conserved, while self-cleavage was not detectable when it was mutated to GAAC (Fig. 5, lanes 1 and 2). These results indicate the absolute requirement of highly conserved GAAAC sequence for self-cleavage, as reported previously (8) (levels of self-cleavage of monomeric transcripts prepared from pPL5 will be analyzed further with content of Table II). To investigate whether self-cleavage could occur *in trans* by double hammerhead structures, complementation experiments were performed. When radioactive GAAAC-transcripts and non-radioactive GAAC-transcripts were snap-cooled and incubated together, only slightly less cleavage was detected when compared to the GAAAC-transcripts alone (Fig. 5, lane 3 versus 1). These results indicate that intramolecular folding is largely favored compared to the intermolecular process. Inversely, when a mixture of radioactive GAAC-transcripts and non-radioactive GAAAC-transcripts were incubated together, only small amounts of self-cleavage was detected, indicating once again that the reaction is largely intramolecular (Fig 5, lane 4).

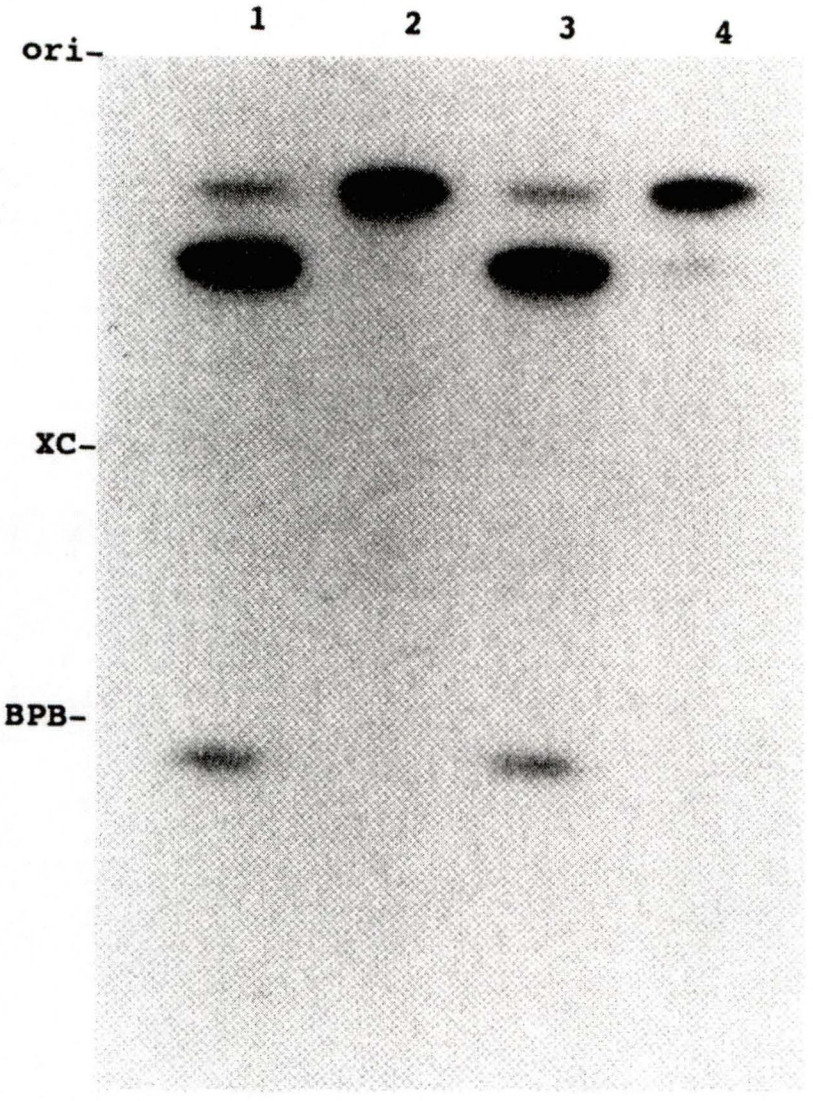


Figure 5. Evaluation of intramolecular and intermolecular self-cleavage of truncated monomeric plus polarity transcripts. In lanes 1 and 2, radioactive transcripts with the GAAAC (control) and GAAC (deletion) sequence, respectively, were independently snap-cooled and incubated at 37°C. In lane 3, radioactive GAAAC-transcripts and non-radioactive GAAC-transcripts were snap-cooled and incubated together. In lane 4, non-radioactive GAAAC-transcripts and radioactive GAAC-transcripts were snap-cooled and incubated together. ori denotes origin of migration and XC xylene cyanol.

The same primers as above were used for PCR amplification with the pPD1 dimeric construction in conditions that allow synthesis of products of both polarities starting from the promoter and truncated either after the first or the second hammerhead site (see Materials and Methods). The smallest templates are identical to the amplification products from pPL5 and exhibit similar self-cleavage during transcription as well as after isolation (data not shown). The largest amplification products allowed to transcribe RNA that possessed two hammerhead structures. The transcripts of both polarities that have the conserved GAAAC sequence at both sites (control) self-cleaved at both sites while those that are mutated to GAAC at the 3' hammerhead sequence self-cleaved only at the 5' site during transcription and after isolation (Table I) (levels of self-cleavage of dimeric transcripts prepared from pPD1 will be analyzed further). If double-hammerhead structures were involved in self-cleavage, the GAAC mutated transcripts would have allowed self-cleavage at the 3' site by action of a non-neighboring GAAAC "catalytic" strand on the 3' site mutated "substrate strand". Hence, the results from PCR-mutated transcripts provide strong evidence for the involvement of only single hammerhead structures in self-cleavage of PLMVd transcripts of both polarities.

Self-cleavage of truncated transcripts

To investigate the importance of the secondary structure and the possibility of interactions with long-ranged sequences influencing the self-cleavage reaction, we prepared various transcripts from pPL5 previously digested with different restriction enzymes and evaluated their autocatalytic efficiency during transcription and after purification (Table II). Self-cleavage efficiency was virtually similar for the transcripts truncated in a region corresponding either to the right arm or the central region of the PLMVd (plus polarity: BamH I, Sty I; minus polarity: EcoR I, Kpn I). The transcripts of minus polarity synthesized after template linearisation with Mbo II, which cuts at the right extremity of the left arm region, self-cleaved at the same level as the full length EcoR I transcripts. Hence,

under the reaction conditions we used, the central and the right arm regions did not influence the self-cleavage efficiency. These results suggest that no important structures and sequences localized in these regions influence self-cleavage. In contrast, self-cleavage was largely improved with transcripts that allow full hammerhead sequence synthesis but are truncated near its extremity, thus preventing the synthesis of the opposite strand of the native structure (minus polarity: Alw26 I; both polarities: PCR-prepared templates). Hence, self-cleavage is likely hindered when adoption of the lowest free energy structure is accomplished. These results propose that the left arm region bearing the hammerhead sequences governs self-cleavage efficiency.

Efficiencies of self-cleavage of the transcripts prepared from the largest amplification products of the pPD1 dimeric construction led to similar conclusions (Table I). In all these transcripts the 5' hammerhead sequences are intact (GAAAC), are not truncated near their extremity, and display a similar self-cleavage to the full monomeric and dimeric transcripts of both polarities (see figure 3 and table II); plus transcripts 5' sites self-cleaved at 60-67% while minus transcripts did so at 51-56% (Table I and see under). In comparison, the 3' hammerhead sequence, which has the GAAAC sequence but is truncated near its extremity, thus preventing the synthesis of the opposite strand of the native structure, exhibits greater self-cleavage efficiencies, like the corresponding monomeric transcripts; plus polarities self-cleaved at 89-93% while minus transcripts at 87-91% (Table I). The improvement of self-cleavage at the 3' hammerhead site has the effect of increasing proportionally the accumulation of one unit-length PLMVd transcripts, which result from the cleavage of both hammerhead sequences. Together these results confirm that self-cleavage is likely hindered when adoption of the lowest free energy structure is realized and that both 5' and 3' hammerhead sequences are independent in their autocatalytic action.

	plus polarity		minus polarity	
	GAAAC	GAAC	GAAAC	GAAC
5'↓ 3'	62/60	63/67	51/54	56/53
5' ↓3'	89/93	-/-	88/91	-/-
5'↓ ↓3'	56/53	-/-	47/43	-/-
5'↓	/65	/62	/53	/54
↓3'	/91	/-	/87	/-

Table I. Self-cleavage efficiency of dimeric transcripts prepared by PCR-amplification from the pPD1 construction. Percentages apply to the hammerhead cleavage sites identified by arrows in the schematic representation of dimers. Percentages are for during transcription (numerator) and after gel purification (denominator). The GAAAC and GAAC columns are for the wild type and mutated 3' hammerhead sequence; the 5' hammerhead is wild type in all cases. - indicates that cleavage was not detected at the arrow(s)-pointed hammerhead sites. The first three rows, in each of the four full dimeric sequences and either during transcription or after purification, are from quantification of bands in a single gel track; so the percentages of cleavage in the cases of a single cleavage come from the addition of the percentage in that case and the one where two cleavages occured. The last two rows, in all cases, bear values that correspond to purified transcripts that had cleaved at one of their sites during transcription. A blank is when the cleavage percentage does not apply.

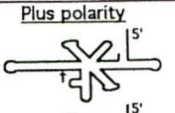
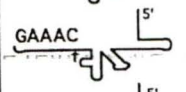
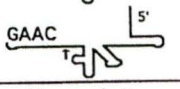
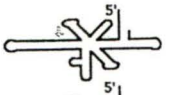
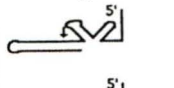
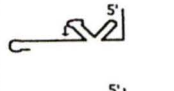
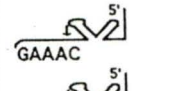
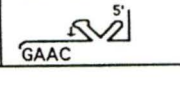
Rest. enz.	Secondary structure	Transcription	Purification
	Plus polarity		
BamH I		69	61
Sty I		68	65
*		94	95
*		—	—
	Minus polarity		
EcoR I		54	52
Kpn I		51	54
Mbo II		54	49
Alw26 I		84	88
*		85	94
*		—	—

Table II. Self-cleavage efficiency in percentage of truncated transcripts during transcription and after purification. Name of restriction enzymes used are reported. (*) is when the templates were prepared by PCR-amplification and GAAAC or GAAC identify the hammerhead sequence. A schematic secondary structure of the truncated transcripts, based on the secondary structure presented at figure 1E, is given to allow localization of their 3' extremity. (-) indicates undetectable level of self-cleavage and the arrow indicates the hammerhead cleavage site.

For all the transcripts studied, the efficiency of self-cleavage is relatively similar either after purification or during transcription (Table I and II) suggesting that the snap-cooling is efficient for the folding of the hammerhead catalytic core. We were anticipating a superior efficiency of self-cleavage during transcription than after purification because the region including the hammerhead sequence is synthesized before its complementary strand, to which it base-pairs, favoring the adoption of the native structure. However in PLMVd, these regions are adjacent and separated only by a few nucleotides (9 nt in plus polarity; 7 nt in minus polarity). We suggest that as soon as the hammerhead sequence is synthesized, the synthesis of the complementary strand is initiated and within a very short time the native secondary structure is adopted. Therefore, for PLMVd, the transcriptional advantage for sequential production and folding of the hammerhead sequences is limited. To further investigate this hypothesis, we performed transcription of a sequence corresponding only to the left arm region (F. Bussière and J.-P. Perreault, unpublished data). The self-cleavage of the resulting transcripts is greatly improved (>95%) when, instead of only one ribonucleotide, two ribonucleotides at a concentration of 10 µM were included in the transcriptional mixture. In these conditions, the polymerase elongation rate is reduced, which probably allowing time for the hammerhead catalytic core to be formed and self-cleavage to occur before that the synthesis of the complementary sequence favors the adoption of the native structure and prevents self-cleavage.

DISCUSSION

The various experiments reported here lead us to conclude that RNAs of both polarities from PLMVd self-cleave by a single hammerhead structure during transcription and also after purification. The secondary structures of the autocatalytic conformation are less stable in terms of free energy than the native rod-like shape structure, which limits the reaction. Like for the satellite RNA of vLTSV (13), the rate-limiting step for PLMVd self-cleavage is the conformational change from the native structure to the alternative structure including the

hammerhead catalytic core, and not the chemical cleavage step. In contrast for several minimal hammerhead structures, Long and Uhlenbeck (14) shown that the chemical cleavage step is rate-limiting. This difference was expected since minimal hammerhead are usually designed to assume mostly the catalytic structure in solution.

Eleven of the fifteen nucleotide differences between the two PLMVd variants, including the one used in the present study and the one used by Hernandez and Flores (1), are in the left arm region which is the most stable region of PLMVd. These nucleotide differences affect locally the native secondary structure (see Fig. 1E), but the free energies (ΔG) of the complete left arm region in rod-like shape of both polarities remains identical, -50 kcal/mol and -53 kcal/mol respectively for plus and minus polarities. Furthermore, these nucleotide differences do not affect the hammerhead secondary structures (Fig. 1C and D); hence, covariation of base-paired nucleotides is observed which suggests selective pressure in favor of the self-cleavage activity. The free energies (ΔG) of the left arm secondary structures including the hammerhead conformation for each variant of both polarities has been estimated by considering two separate domains: the sequence forming the hammerhead structure and the balance of the sequence of the left arm. The estimated free energies (ΔG) of the left arm region in hammerhead structure reveal that the variant used here is slighty more stable than the variant characterized by Hernandez and Flores (1). Consequently, the free energy difference (ΔG) of the native structure and that of the hammerhead structure is smaller (~20%) for the variant used in the present study. This difference could explain, at least partially, why the variant we used self-cleaved more efficiently after purification (> 50%) than the variant used by Hernandez and Flores (~10% in the absence of formamide or snap-cooling on dry ice; ref. 1). Furthermore, when we estimated the free energy of the structures adopted by the plus polarity hammerhead sequence of the variant used by Hernandez and Flores, the autocatalytic structure was not the most stable alternative structure with a free energy of ~20% less. This difference may also account for the lower level of self-cleavage of their variant. Our results support the hypothesis that the efficiency

of self-cleavage is governed by the stability of the structures including the hammerhead motif as compared to the native structure. However, these results do not explain all the differences of self-cleavage efficiency between variants of both polarities and do not account for the putative contribution of any tertiary interactions. Efficient adoption of alternative structures by viroids, including the hammerhead conformation, is of crucial importance for their replication since the autocatalytic cleavage releases monomeric copies from multimeric strands. Study of viroid conformational isomers is obviously important to understand related biological activities.

In contrast to ASBVd, PLMVd shows the involvement of only single hammerhead structures for the processing of multimeric strands. This characteristic was previously associated only with satellite RNAs. Our study indicates that the mode of self-cleavage is not a strict classification character in order to differentiate between viroids and other related satellite RNAs. On the other hand, it supports the proposed phylogenetic position of PLMVd between ASBVd and vLTSV (1). ASBVd has been proposed as an the evolutionary link between viroids and satellite RNAs in a monophylogenetic tree (15). PLMVd appears to evolve from vLTSV and it has conserved its self-cleavage properties. Therefore PLMVd may be a unique case among viroids.

ACKNOWLEDGEMENT

We thank Alain Lavigueur for useful comments concerning the manuscript. This work is sponsored by an establishment fund from Fonds de Recherche en Santé du Québec (FRSQ) and a grant from Natural Sciences and Engineering Research Council of Canada (NSERC) to J.-P. P. F.L. and F.B. have a predoctoral fellowship from the Faculté de Médecine de l'Université de Sherbrooke. J.-P. P. is a Junior Scholarship of FRSQ.

REFERENCES

1. Hernandez,C. and Flores,R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 3711-3715.
2. Symons,R.H. (1990) *Seminars in virology*, **1**, 75-81.
3. Symons,R.H. (1992) *Ann. Rev. Biochem.* **61**, 641-671.
4. Branch,A.D. and Robertson,H.D. (1984) *Science* **223**, 450-455.
5. Bratty,J., Chartrand,P., Ferbeyre,G. and Cedergren,R. (1993) *Biochim. Biophy. Acta* **1216**, 345-359.
6. Tsagris,M., Tabler,M. and Sänger,H.L. (1991) *Nucleic Acids Res.* **19**, 1605-1612.
7. Diener,T.O. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9370-9374.
8. Forster,A.C., Davies,C., Sheldon,C.C., Jeffries,A.C. and Symons,R.H. (1988) *Nature* **334**, 265-267.
9. Davies,C., Sheldon,C.C. and Symons,R.H. (1991) *Nucleic Acids Res.* **19**, 1893-1898.
10. Epstein,L.M. and Pabon-Pena,L.M. (1991) *Nucleic Acids Res.* **19**, 1699-1705.
11. Flores,R., Hernandez,C., Desvignes,J.C. and Llàcer,G. (1990) *Res. Virol.* **141**, 109-118.
12. Zuker,M. (1989) *Science* **244**, 48-52.
13. Sheldon,C.C. and Symons,R.H. (1989) *Nucleic Acids Res.* **17**,5665-5676.
14. Long,D.M. and Uhlenbeck,O.C. (1994) *Proc. Natl. Acad. Sci. USA* **91**,6977-6981.
15. Elena,S.F., Dopazo,J., Flores,R., Diener,T.O. and Moya,A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 5631-5634.

Description du travail

L'objectif principal de cet article était de caractériser l'activité d'autocoupure chez PLMVd *in vitro*. En utilisant différents transcrits correspondant au viroïde ainsi que certains mutants inactifs des motifs catalytiques, nous avons montré que la coupure des ARN multimériques sur les deux polarités dépend de l'adoption d'une structure en simple *hammerhead*. La réaction est excessivement rapide (<15 sec), ce qui suggère que le repliement adéquat du motif catalytique constitue l'étape limitante de la réaction. L'efficacité de l'autocoupure varie de 55% à 70%, dépendamment de la polarité, et semble dépendre de la différence de stabilité entre la structure catalytique active et la structure native double brin de la région qui contient le motif catalytique. En effet cette structure double brin inhibe la réaction d'autocoupure en empêchant la formation du *hammerhead*. L'adoption transitoire d'une structure catalytique active durant la transcription permet au viroïde de promouvoir la coupure des brins multimériques.

Implication dans le travail

J'ai principalement contribué à la réalisation des expériences sur la formation transitoire du *hammerhead* durant la transcription. J'ai également participé aux discussions entourant le déroulement des travaux. J'ai, entre autres, suggéré la construction des mutants afin de prouver sans équivoque que l'autocoupure dépend d'une réaction en simple *hammerhead*.

CHAPITRE 2:

Localisation cellulaire et réplication de PLMVd

2.1 ARTICLE:

Subcellular localization and the rolling-circle replication of peach latent mosaic viroid: hallmarks of group A viroids

F. Bussière, J. Lehoux, D. A. Thompson,

L. J. Skrzeczkowski et J.-P. Perreault

Département de Biochimie

Université de Sherbrooke

Sherbrooke, Québec, J1H 5N4, Canada

Article publié dans: *Journal of Virology*, Vol. 73, 6353-6360, 1999.

ABSTRACT

Peach latent mosaic viroid (PLMVd) is an RNA species in which both the plus and minus strands possess hammerhead structures essential for replication via a rolling circle mechanism. We report here the use of various hybridization techniques to characterize PLMVd from infected peach leaves. Using *in situ* hybridization we show at a tissue level that PLMVd strands of both polarities concentrate in the cells forming the palisade parenchyma. At a cellular level, PLMVd was found to predominately, if not exclusively, accumulate in the chloroplasts. Northern blot analyses demonstrated that this viroid replicates via a symmetric mode involving the accumulation of both circular and linear monomeric strands of both polarities, with no multimeric conformers being detected. Finally, dot blot hybridizations revealed that PLMVd strands of both polarities accumulate equally , but that the relative concentrations vary by more than 50 fold between peach cultivars. These results confirm two new hallmarks for the classification of viroids. Group A viroids (e.g. PLMVd), which possess hammerhead structures, are chloroplastic and replicate via the symmetric mode. In contrast, group B viroids, which share a conserved central region, are nuclear and replicate via an asymmetric mechanism. The implications of these fundamental differences are discussed.

INTRODUCTION

Viroids are small (~300 nt), single-stranded, circular RNAs that infect higher plants causing significant losses in the agricultural industry (1, 2;reviews). The 26 known viroid species have been classified in two groups, A and B (1, 3; reviews). This classification is based primarily on whether or not a viroid possesses the five structural domains characteristic of a group B viroid. The group B viroids are further subdivided on the basis of both the sequence and the length of a highly conserved central region (CCR). Three viroids possess no sequence or structural similarity with the group B viroids, and have been classified in group A. These latter viroids possess self-cleaving hammerhead motifs that are essential for their replication (see below). This classification is supported by phylogenetic reconstructions in which a group A viroid (ASBVd; avocado sunblotch viroid) has been proposed as an evolutionary link between the classical group B viroids and the plant viroid-like satellite RNAs (4).

In infected cells, viroids replicate in a DNA-independent manner via a rolling circle mechanism that follows either a symmetric or an asymmetric mode (1; Fig. 1). In the symmetric mode, the infecting circular monomer (which is assigned plus polarity by convention) is replicated into linear multimeric minus strands which are then spliced and ligated yielding minus circular monomers. Using the latter RNA as template, the same three steps are repeated so as to produce the progeny. In contrast, in the asymmetric mode the linear multimeric strands serve directly as the template for the synthesis of linear multimeric plus strands. Therefore, both the linear and circular minus monomers are produced only within the symmetric mode. For example, the fact that minus circular monomeric strands of ASBVd are present in RNA isolated from infected avocado plants is taken as evidence that ASBVd replicates via the symmetric mode (5, 6). Similarly, the fact that the circular minus monomer of potato spindle tuber viroid (PSTVd) has not been found in plants infected by this viroid has been taken as evidence that it replicates via the asymmetric mode (7).

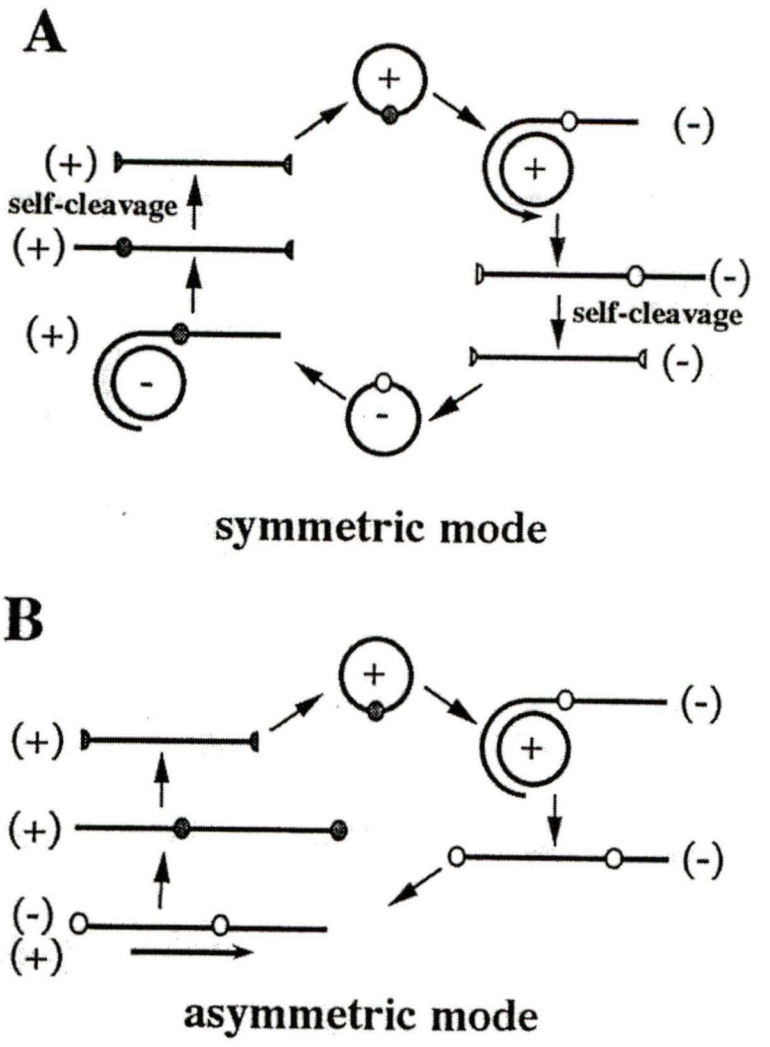


Figure 1. Schematic representation of the mechanism of viroid rolling circle replication following either the symmetric (A) or the asymmetric (B) mode. The polarity of the strands is indicated in the parenthesis, and the small circle on the strands denotes the cleavage site. The process begins with the infecting circular (+) strand.

To date most of our knowledge of viroid biology comes from studies of group B viroids (e.g. , PSTVd). These viroids accumulate in the nucleus of infected cells (Schumacher *et al.*, 1983; Harders *et al.*, 1989; Bonfiglioli *et al.*, 1996), where they are believed to replicate via the asymmetric mode using host RNA polymerase II as the replicase enzyme. The actual mechanism of the processing of the resulting multimeric strands is still a matter of debate. It was proposed that a host endoribonuclease catalyzes the cleavage of multimeric strands into monomers (revised in Flores *et al.*, 1997); however, a recent report has shown that coconut cadang cadang viroid processing could be mediated by a new self-cleaving motif under specific conditions (Liu and Symons, 1998). In contrast, studies on ASBVd have shown that this group A viroid is located mainly in the host's chloroplasts (Bonfiglioli *et al.*, 1994; Lima *et al.*, 1994), and that none of the three nuclear RNA polymerases participate in its replication (Flores *et al.*, 1997). In this system the processing of the multimeric intermediates is mediated by self-catalytic hammerhead motifs. Therefore, it has been proposed that both the replication mode and the viroid subcellular localization may be linked, and potentially constitute a fundamental difference between the two major groups of viroids (Flores *et al.*, 1997). In order to confirm this analysis it is essential to determine whether or not the ASBVd features are common to other group A viroids, more specifically to a member of the PLMVd-sub-group (see below).

PLMVd is an RNA species of 335-338 nucleotides (nt) which causes the latent mosaic of peach trees (Hernandez and Flores, 1992). Both the plus and minus multimeric PLMVd strands have been shown to efficiently self-cleave *in vitro* using hammerhead structures (Hernandez and Flores, 1992; Beaudry *et al.*, 1995). Due to the presence of self-cleavage properties, and the absence of a known CCR, PLMVd was proposed to be a group A viroid (Hernandez and Flores, 1992). PLMVd, ASBVd and chrysanthenum chlorotic mottle viroid (CChMVd) are the only hammerhead containing species for which the group A classification is confirmed. Unlike other viroids that fold into rod-like or quasi-rod-like structures, CChMVd and PLMVd adopt an unusual branched secondary structure which

makes them the only viroids that are insoluble in 2 M lithium chloride (Navarro and Flores, 1997). These results prompted the classification of the self-cleaving group A viroids into two subgroups, one formed by ASBVd and one by PLMVd and CChMVd (Navarro and Flores, 1997). In this report, we characterize the PLMVd replication intermediates which accumulate in infected peach leaves, and determine the tissue and subcellular localization of this RNA species in order to strengthen the existing criteria for the classification of viroids.

Results

In situ hybridization of PLMVd strands

In order to determine the localization of PLMVd in tissue, leaf pieces from both healthy and PLMVd-infected Siberian C peach plants were harvested, fixed, and studied by *in situ* hybridization with digoxigenin-11-UTP (DIG-UTP) labeled riboprobes (Fig. 2). The PLMVd infection was confirmed by Northern blot hybridization (see below). In a first set of experiments, the detection of the DIG-labeled hybrids was performed with alkaline phosphatase conjugated anti-DIG antibodies (e.g. Fig. 2A-E). No hybridization signal was detected when PLMVd-infected samples were incubated in the absence of a probe or with a DIG-PSTVd probe (Fig. 2A-B). Similarly, no DIG-labeled hybrids were detected in healthy peach leaves (Fig. 2C). In contrast, riboprobes of both polarities of PLMVd reacted with PLMVd-infected samples (Fig. 2D-E). These results were consistent within the various cultivars examined, with the hybridization signals being limited to PLMVd-infected peach-leaves (data not shown). At a tissue level, the DIG-labeled hybrids were observed to concentrate in the cells forming the palisade parenchyma. Several DIG-labeled hybrids appeared in each cell; unfortunately, the low resolution of light microscopy could not discern the specific subcellular localization. Variations in the hybridization conditions, washings and DIG-detection revealed different amounts of DIG-labeled hybrids, but in all cases the proportion of the plus and minus strands remained identical.

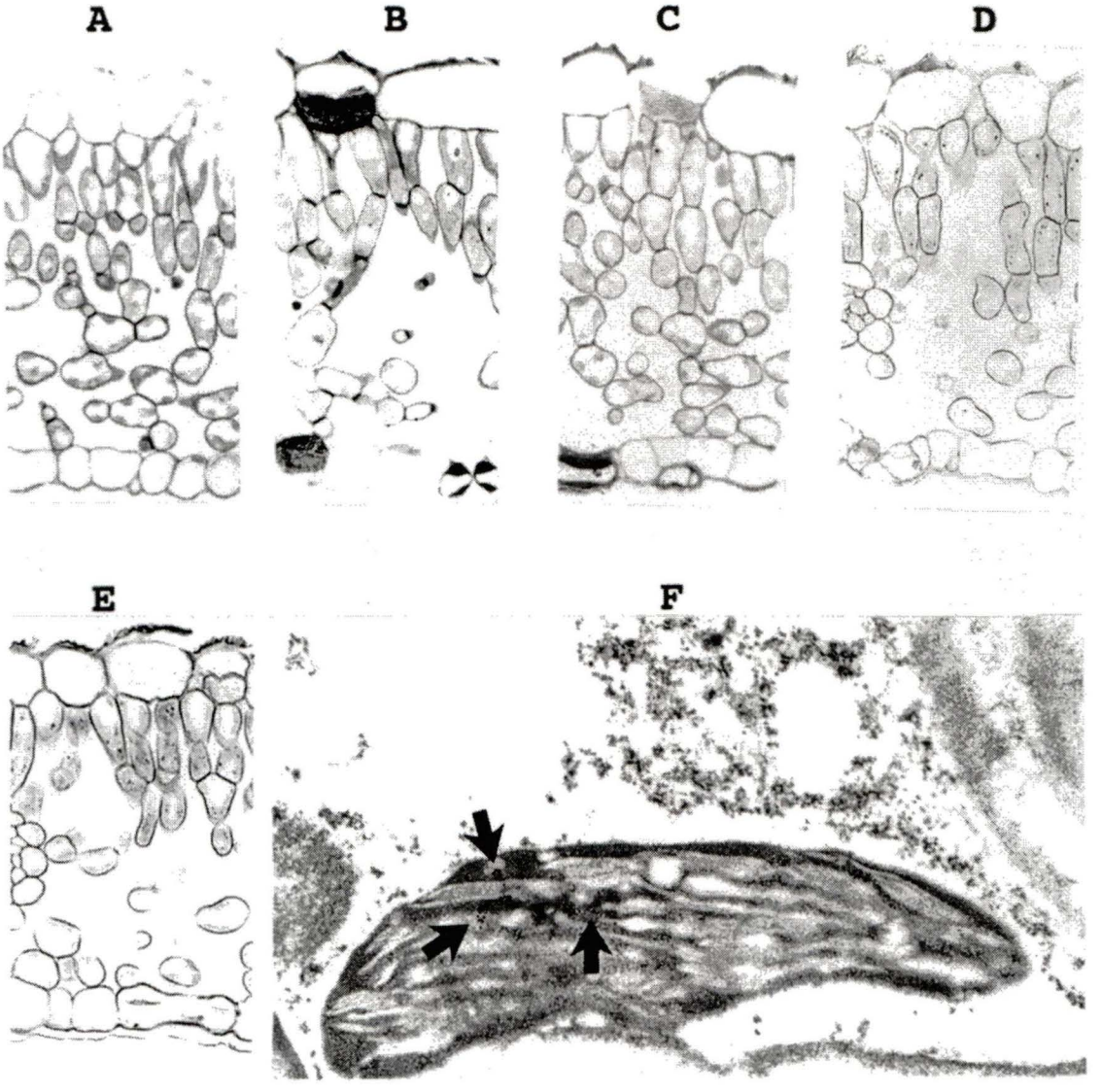


Figure 2. Representative light microcope in situ hybridizations using DIG-labeled riboprobes (A-E) Observations of the Dig-labeled hybrids by light microscope. (A) and (B) Sections of PLMVd-infected leaves hybridized in either the absence or the presence of DIG-PSTVd riboprobe, respectively. (C) Section of a healthy peach-leaf probed with a minus polarity PLMVd-riboprobe. (D) and (E) Sections of PLMVd-infected peach leaves probed with either the plus or minus PLMVd-riboprobes, respectively. Control panels (A-C) were overstained to ensure the detection of trace amounts of PLMVd strands. (F) Typical electron micrograph of the hybridization of PLMVd-infected peach leaves with the minus strand PLMVd-riboprobe. The arrows point to clusters of grains representing PLMVd accumulated strands.

In a second set of experiments the target-probe hybrids were revealed using anti-DIG antibodies coupled to gold particles, and were visualized by electron microscopy in order to determine the subcellular localization of PLMVd. Figure 2F shows a typical result when probing with the PLMVd riboprobe of minus polarity. *In situ* hybridization revealed clusters of PLMVd strands of both polarities in the chloroplasts of infected leaves. In contrast, only negligible amounts of gold particles were found in healthy peach leaves due to the non-specific binding of the labeled antibody (data not shown). Statistical analysis involved counting the colloidal gold particles in 15 to 20 cells per experiment. In some experiments more than a thousand particles were counted in PLMVd infected samples. These analyses showed an increase of at least 8 fold (with an average of 11.3 fold) in the number of gold particles detected in the chloroplasts of PLMVd-infected peach leaves as compared to healthy ones. No corresponding increase was observed in other organelles. Similar increases in chloroplast labeling was observed with both plus and minus PLMVd-riboprobes on PLMVd infected samples, but not PSTVd-riboprobes (data not shown). These evaluations revealed that more than 80 % (81-96%) of the total grains observed in the cell were located in the choloroplasts. The chloroplasts in PLMVd-infected tissues contained an average of 10-14 grains depending the experiment. The remaining grains appeared in the cytoplasm, the vacuoles and the nuclei, while other cellular structures such as mitochondria and the cell wall had negligible counts. Finally, approximately the same amount of grains were detected irrespective of the polarity of the PLMVd-riboprobe used, suggesting that both PLMVd strands accumulate to the same degree. These results clearly show that PLMVd is predominately located in the chloroplasts.

Detection of PLMVd replication intermediates

In order to identify the PLMVd replication intermediates which accumulate in cells, RNA samples from both healthy and PLMVd-infected leaves were isolated and analyzed by Northern blot hybridization. Because PLMVd strands of both polarities have the ability to self-cleave in the presence of magnesium, RNA samples were isolated according to various procedures (see the Materials and methods section). Two of the four methods used included 5 mM EDTA in their extraction buffers so as to chelate any cations and thereby prevent any self-cleavage of the multimeric strands. All RNA samples were electrophoresed through both agarose and polyacrylamide gels, transfered to filters and successively probed with plus and minus PLMVd-riboprobes (Fig. 3).

Initially, the RNA samples were analyzed on 1.3% agarose gels in order to verify whether or not PLMVd multimeric strands were accumulated. Nucleic acids were hybridized with radioactive PLMVd-riboprobes synthesized by run-off transcription (see the Materials and methods section). Figure 3A shows a typical autoradiogram of a hybridization performed using the plus riboprobe to detect minus polarity PLMVd strands. Lanes 1 and 2 are controls containing synthetic PLMVd transcripts of plus and minus polarities, respectively. Cross-hybridization between the probe and RNA strands of the same polarity was detectable only after over-exposure of the gel (i.e. < 2%). Irrespective of the RNA isolation procedure used, a band corresponding to the PLMVd monomers was consistently detected in the infected samples, but not in the healthy ones (compare lanes 4-10 to lane 3). The position of this band varied slightly depending the RNA isolation procedure; most likely due to the various RNA isolation methods yielding RNA samples which contain different salts in varying concentrations. In addition, the intensity of the band varied depending the extraction procedure used and on the particular cultivar from which the sample was isolated. The RNA nature of the species was confirmed by its resistance to DNase treatment (lane 11) and susceptibility to alkaline hydrolysis (lane 12). An additional non-specific RNA band of ~ 1 kb was detected in both infected and healthy samples, and

appeared to be more abundant when the Qiagen RNA extraction procedure, which favors the isolation of longer RNA species, was used. Regardless of the RNA extraction method used, over-exposure of the Northern blots resulted in the appearance of this band (data not shown). However, since the greatest amount was observed in samples from healthy leaves (lane 3), we believe that it is not a PLMVd replication intermediate.

Both the plus and minus PLMVd-riboprobes were used to analyze the PLMVd replication intermediates in several peach cultivars. The minus PLMVd-riboprobe (Fig. 3B) detected a band corresponding to monomer plus strand PLMVd (the non-specific band of ~1 kb was detected in one sample), while probing of the same filter (after being stripped) with the plus PLMVd riboprobe detected only monomeric PLMVd of minus polarity (Fig. 3C). These results suggest that both strands of PLMVd are accumulated as monomers.

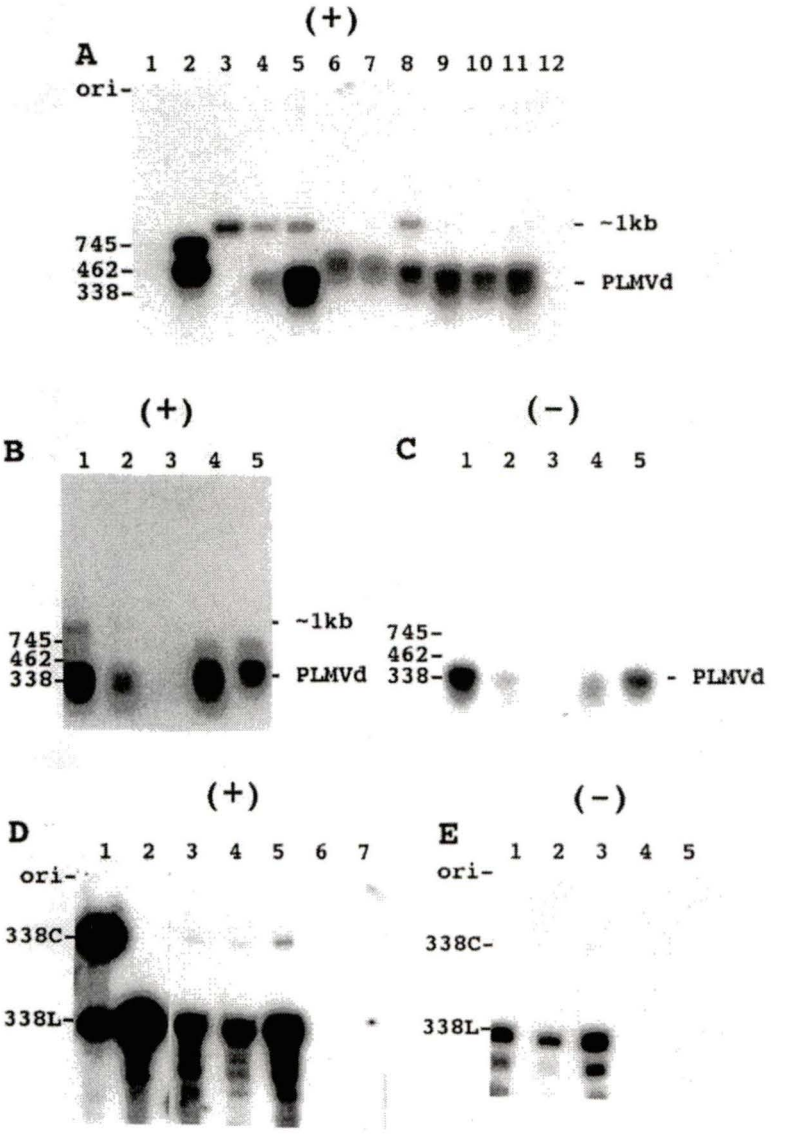

(+)
A
1 2 3 4 5 6 7 8 9 10 11 12
ori-
745-
462-
338-
- ~1kb
- PLMVd
(+)
B
1 2 3 4 5
745-
462-
338-
- ~1kb
- PLMVd
(-)
C
1 2 3 4 5
745-
462-
338-
- PLMVd
(+)
D
1 2 3 4 5 6 7
ori-
338C-
338L-
(-)
E
1 2 3 4 5
ori-
338C-
338L-

Figure 3 (page précédente). Autoradiograms of Northern blot hybridizations of RNA samples isolated from both healthy and PLMVd-infected peach leaves. RNA samples were fractionated on either agarose (**A-C**) or polyacrylamide (**D** and **E**) gels and then blotted onto nylon filters. The polarity of the PLMVd strands detected is indicated by the symbol (+) and (-) at the top of the panel. (**A**) RNA samples were isolated according to various extraction procedures: RNeasy kit in presence of EDTA (Qiagen) (lanes 3 to 5); Tris-EDTA isolation (lanes 6 and 7); Rneasy Plant mini kit (Qiagen) in the absence of additional EDTA (lane 8); and the PEG-precipitation procedure (lanes 9 to 12). In lanes 1 and 2, non-radioactive synthetic PLMVd transcripts of plus (761 nt and 588 nt) and minus (745 nt and 462 nt) polarity, respectively, were loaded as controls. Lane 3 is a sample of healthy GF-305 peach cultivar; Lanes 4, 6, 8, 9 11 and 12 contain RNA samples isolated from a PLMVd-infected RedGold peach cultivar. The samples in lanes 11 and 12 were subjected to either DNase treatment, or alkaline hydrolysis prior electrophoresis. Lanes 5 and 7 correspond to samples from a PLMVd-infected Redhaven cultivar, while lane 10 is from the leaves of the Agua cultivar. Panels (**B**) and (**C**) show the same filter following with either the plus or minus polarity riboprobe, respectively. All samples were isolated by the PEG-precipitation procedure. Lanes 1 and 2 were isolated from different leaves of a Redhaven peach. Lane 3 is an RNA sample from a healthy Bailey cultivar. Lanes 4 and 5 are RNA samples of PLMVd-infected Redhaven and Siberian C cultivars. Panels (**D**) (lanes 3-7) and (**E**) (lanes 1-5) show a filter following PAGE and blotting that was probed with both the plus and minus polarity riboprobes, respectively. In panel D lanes 1 and 2 are synthetic linear and circular PLMVd controls, while lanes 3-5 (which correspond to lanes 1-3 of panel **E**) are from PLMVd-infected Redgold, Hardired and Siberian C cultivars. Lanes 6 and 7 (which correspond to lanes 4 and 5 of panel **E**) show samples from healthy GF-305 and Baily cultivars. Adjacent to the gel the positions of several PLMVd transcripts are indicated as size references including both the circular (338C) and linear (338L) strands, and the mixture of circular and linear molecules

(PLMVd). The position of the unknown ~1 kb species is also indicated in panels (**A**) and (**B**).

Regardless of the extraction method used, no multimeric strands of either polarity were detected except after over-exposure. The presence of 5 mM EDTA in the buffers, which chelates the bivalent cations necessary for self-cleavage, did not increase the concentration of multimeric strands. To rule out the possibility that both the linear multimers and circular monomers self-cleaved to yield mainly linear monomers, we performed a set of control extractions. Synthetic radiolabeled PLMVd transcripts were synthesized as described previously (i.e. linear dimers and circular monomers; Côté and Perreault, 1997), and added to the ground leaf powder. Ribonucleic acids of these mixtures were extracted using the RNeasy plant mini kit (Qiagen) either with or without 5 mM EDTA, and were subsequently fractionated on polyacrylamide gels. Radiolabeled transcripts were revealed by autoradiography (data not shown). Independent of the presence of EDTA, only trace amounts of linear monomer strands were detected, suggesting that self-cleavage was limited during these extractions. Therefore, self-cleavage during extraction is not responsible for either the absence of multimeric strands, or for the low concentration of circular conformers observed.

Polyacrylamide gel electrophoresis (PAGE) followed by Northern blot hybridization was used to resolve circular and linear PLMVd conformers (Fig. 3D and E). Only the two bands corresponding to the linear and circular PLMVd monomers were detected in RNA samples isolated from infected peach leaves. Multimeric strands of either polarity were not observed. The proportion of linear versus circular monomeric molecules was estimated for both the plus and minus strands of several PLMVd cultivars (Table I). This proportion varied from 8 to 16 (average = 11) linear conformers per circular molecule for the plus strands, and from 12 to 16 (average = 14) linear conformers per circular molecule for the minus strand. These results show that PLMVd strands of both polarities are predominantly accumulated as linear conformers.

TABLE I. Quantitative analysis of PLMVd in peach tissue.[ab]

peach cultivar	infected plants	linear vs circular strands		quantity per mg tissues	
		plus	minus	plus	minus
Hardired	+	13:1	13:1	0.4	0.3
Siberian C	+	12:1	16:1	2.2	2.0
Redhaven	+	9:1	14:1	5.5	3.2
Redgold	+	16:1	14:1	0.1	0.2
Agua	+	8:1	12:1	0.4	0.2
Harrow Beauty	+	8:1	15:1	1.5	1.2
Bailey[c]	-	--	--	--	--
Siberian C[c]	-	--	--	--	--
GF-305[c]	-	--	--	--	--

[a]The proportions of linear:circular PLMVd strands of both polarities were estimated by the quantification of Northern blot hybridizations.
[b]The quantity PLMVd in ng per mg tissue was determined by dot blot hybridization.
[c]Healthy peach leaves

Quantity of PLMVd strands

In order to establish the quantity (ng PLMVd/ mg tissues) of PLMVd strands, we performed dot blot hybridizations. This technique does not differentiate between the linear and circular conformers; however, since we know from the Northern blot hybridizations the linear conformer predominates this is not a concern. Only RNA samples isolated with the PEG-precipitation including the 3.5 M LiCl known to favor PLMVd extraction (Navarro and Flores, 1997), were used. Several known concentrations of linear monomeric (338 nt) PLMVd transcripts of both polarities were used as standards. Serial dilutions of RNA from both healthy and infected leaves were spotted onto filters and analyzed with PLMVd riboprobes of both polarities. The results are reported in Table II. PLMVd was only detected in RNA samples from infected leaves, and not in those from healthy ones (< 0.01 ng / mg tissue). The concentrations of both the plus and minus PLMVd strands was relatively constant within each cultivar, except in the case of the Redgold cultivar in which slightly more plus strands were detected. In contrast, the concentrations varied significantly between cultivars, ranging from 0.1 to 5.5 ng of PLMVd strands per mg of tissue. PLMVd was more concentrated in Siberian C, Redhaven and Harrow Beauty and less concentrated in Hardired, Redgold and Agua peach trees. The observation that the strands of both polarities accumulate at approximately similar concentrations within a cultivar, but that these concentrations varied between the cultivars correlates with the Northern blot analyses. Total PLMVd strands accumulate to levels comparable to PSTVd, which has been shown to accumulate to a concentration of 1 ng per mg of tissue (Colpan *et al.*, 1983).

Discussion

Subcellular localization is a trademark of a viroid group

Using *in situ* hybridization we have shown that PLMVd replicates predominantly in the the palisade parenchyma cells of infected peach leaves (Fig. 2). However, it remains unknown if this cellular localization is a result of the high metabolic activity occurring in these cells, or of a PLMVd molecular determinant to either enter or replicate primarily in these cells. Within these cells we show that PLMVd strands of both polarities accumulate in the chloroplasts (> 80%). The similar localization of PLMVd and ASBVd, even though they are members of different subgroups (i.e. PLMVd- and ASBVd-sub-groups, respectively), suggests that the subcellular location of a viroid is a fundamental characteristic of their group classification (Fig. 4A). Group A viroid are found in chloroplasts, while group B viroids are found in nuclei (see Introduction). Clearly this fundamental difference has important implications in the molecular biology and evolutionary origin of both viroid groups.

Because of the difference in subcellular location, nuclear and chloroplastic viroids are likely to use different mechanisms to ensure their replication. For example, host nuclear RNA polymerase II has been proposed to replicate group B viroids (reviewed in Flores *et al.*, 1997), but has been shown to be unable to replicate both ASBVd (Marcos and Flores, 1992) and PLMVd (F. Lareau, 1996). In fact, the host polymerase supporting the replication of group A viroids has yet to be identified. Indeed, two different RNA polymerases have been shown to transcribe chloroplastic genes: one is a bacteriophage-like RNA polymerase, while the other appears to be related to *Escherichia coli* RNA polymerase (reviewed in Link, 1996). One of these RNA polymerases likely replicates group A viroids.

Viroids of both groups also use different strategies in the processing of their multimeric strands. This step is mediated by the hammerhead self-catalytic motif in group A viroids, while group B members are thought to use either a host endoribonuclease or some other catalytic motif (see the Introduction). The use of a self-cleavage motif by the chloroplastic viroids might be the result of the absence of an appropriate endoribonuclease in the chloroplast. This hypothesis is supported by the lack of enzyme-mediated processing of RNA (e.g. intron) transcribed in chloroplasts. In contrast, plant viroid-like satellite RNAs are most likely replicated in the cytoplasm by their helper virus replicase. The cytoplasm may also lack an appropriate specific endoribonuclease; hence, small satellite RNAs, which are believed to replicate by a rolling circle mechanism, use either hammerhead self-cleavage or have evolved other catalytic motifs (e.g. hairpin self-catalytic motif) for the cleavage of their multimeric strands. If we agree with the hypothesis of a monophyletic origin for both the group A and group B viroids, and for the related satellite RNAs (Elena et al., 1991; see Fig. 4A), an essential factor for their divergence over time is their cellular localization. The recent observation that the group B viroids may use a new self-cleavage motif (Liu and Symons, 1998) supports the belief that the ancestor of today's viroids possessed a catalytic motif which diverged depending on cellular location. If satellite RNA and viroids share a common ancestor, the ancestral catalytic motif of both cytoplasmic and chloroplastic small plant RNA pathogens would be a hammerhead.

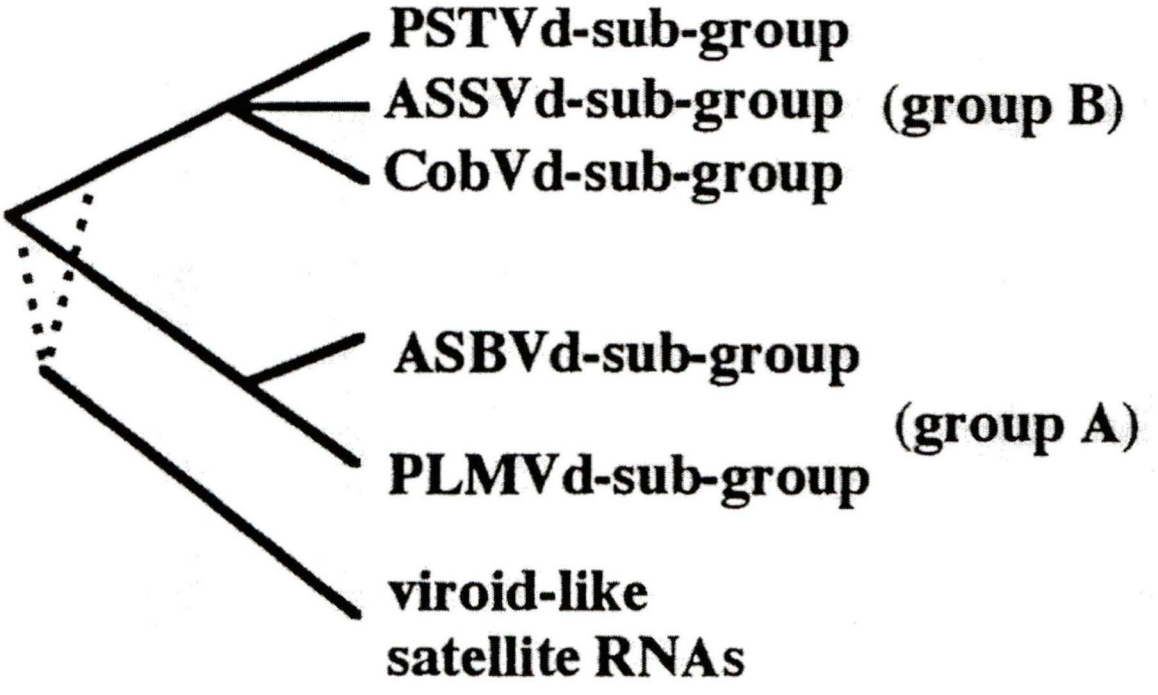


Figure 4. Schematic phylogenic reconstruction of viroids and viroid-like satellite RNAs based on physical characteristics. The dotted lines for the satellite RNAs indicate that their branching remains to be determined (see text).

The PLMVd rolling circle mechanism

The recent discovery of retroviroid-like elements such as rCarSAV (carnation stunt associated viroid; Daros and Flores, 1995) prompted us to investigate whether or not PLMVd was integrated into the host genome. Southern blot hybridizations and PCR amplifications using various primers and DNA from infected peach-leaves as template suggests the lack of a PLMVd counterpart in either the host genome or the extrachromosomal DNA (unpublished data; S. Laurandeau, F. Bussière and J.-P. Perreault). Thus, the replication cycle of PLMVd depends solely on RNA intermediates, as is observed for all other viroids.

Northern blot hybridizations demonstrate that PLMVd replication follows a symmetric pathway involving predominantly, if not exclusively, the accumulation of circular and linear monomeric strands of both polarities (summarized in Fig. 4B). Clearly, the linear conformation is the most abundant form, accumulating to levels at least eight times higher than the circular molecules do (Table I). This pattern has also been observed in the newly discovered group A viroid CChMVd (Navarro and Flores, 1997). In addition, a small, circular viroid-like RNA (csc RNA 1) associated with a cherry disease follows the same pattern except that it seemed to accumulate circular and linear conformers at the same level (Di Serio *et al.*, 1997). PLMVd, CChMVd and csc RNA 1 appear to self-cleave to completion *in vivo*, although their *in vitro* self-cleavage efficacy only ranges from 10 to 60% (Beaudry *et al.*, 1995; Navarro and Flores, 1997; Di Serio *et al.*, 1997). The lack of accumulated PLMVd multimeric strands in infected peach leaves is supported by earlier *in vitro* experiments. When PLMVd was transcribed *in vitro* under conditions of slow polymerase activity, which favours sequential production and folding of the hammerhead sequences, self-cleavage of the resulting transcripts was greatly increased (>95%) as compared to the standard conditions (50-60%) (Beaudry *et al.*, 1995). The ability of viroids that possess autocatalytic sequences to self-cleave appears to depend on the capacity of the RNA strands to adopt a conformation different from their most stable native structure

(Beaudry *et al.*, 1995). Polymerase processivity or an unknown molecular adaptation (e.g. a cofactor, interacting with a cellular component mediating structural changes, etc.) may explain the high level of self-cleavage of these RNAs *in vivo*. RNA species other than PLMVd, CChMVd and csc RNA 1 possessing hammerhead structures on strands of both polarities were shown to differ by the accumulation of a set of plus multimeric strands (1- to 11-mer in length) and smaller minus multimeric strands (1- to 3-mer in length) (Sheldon and Symons, 1993; Silver *et al.*, 1994; Hutchins *et al.*, 1985).

In summary, the three group A viroids (i.e. PLMVd, ASBVd and CChMVd) replicate in infected cells through the symmetric mode involving self-cleavage activity. Thus, symmetric replication may be thought of as a second hallmark of the group A viroids. Furthermore, since ASBVd accumulates multimeric strands, while PLMVd and CChMVd do not, group A viroids can be classified into the ASBVd- or PLMVd-subgroups using this feature.

Predomidantly accumulated PLMVd strands

The concentration of PLMVd linear and circular monomeric strands detected by dot blot analysis of RNA isolated from infected peach leaves was almost identical (Table II). Using a different approach, a similar observation have been reported for csc RNA 1 (Di Serio *et al.*, 1997). Northern blot analysis of CChMVd replication intermediates suggests that plus polarity strands are slightly more abundant than their minus counterparts (Navarro and Flores, 1997). These three RNA species are the only ones known to accumulate monomeric linear and circular strands of both polarities to approximately the same level. Therefore, the rolling circle replication of these species is a perfectly symmetrical mechanism using identical intermediates for both two polarities. Such symmetry implies that the plus and minus strands are utilized equally by the replication machinery to produce multimeric conformers which subsequently self-cleave and are ligated to the same extent. Therefore, the molecular components of the viroid responsible for enzyme recognition and ligation of

the strands of both polarities is probably the same, as is observed for hammerhead self-cleaving motifs.

The hybridization techniques used here have yielded valuable information on both the cellular localization and the rolling circle replication of PLMVd. However, several questions remain unanswered. For example, what is the stability and the role of the monomeric linear conformer? The high levels of PLMVd linear monomers suggests that this conformation is as stable as the circular conformers, even though the latter are throught to be stabilized by their absence of extremities. To prevent degradation from host exonucleases, both the 5' and 3' extremities of the linear conformers must be embedded within the RNA structure. This hypothesis is supported by the observation that PLMVd monomeric strands are difficult to end-label. Specifically, monomeric PLMVd strands of both polarities produced by self-cleavage from multimeric strands are relatively inefficiently labeled with ^{32}P at both the 5'-end by T4 polynucleotide kinase in the presence of [g-^{32}P]ATP, and at the 3'-end by T4 RNA ligase in the presence of the p^{32}Cp after prior dephosphorylation (J.-P. Perreault, unpublished data). Thus, it seems reasonable to suspect that PLMVd linear monomers can accumulate because their extremities are located within the RNA structure and are protected from the host exonuclease. However, it remains unknown why these linear monomers are accumulated. One possible explanation is an inefficient circularization step. Self-ligation has been demonstrated for the circularization of PLMVd strands *in vitro*, but remains to be proved *in vivo* (Lafontaine *et al.*, 1995; Côté and Perreault, 1997). *In vitro* self-ligation is relatively inefficient (i.e. ~10%), and thus favors the accumulation of linear monomers. This correlates with the ratio of linear versus circular PLMVd strands estimated in this report (11:1 and 14:1 for the plus and minus polarities, respectively). A second possibility is that the linear monomers play a direct role in the life cycle of this infectious RNA species.

In conclusion, PLMVd strands of both polarities accumulate primarily in the chloroplasts of leaves, and this viroid replicates by a rolling circle mechanism operating in the symmetric mode. These two characteristics are hallmarks for inclusion in the group A viroids. However, symmetric or asymmetric replication is related to the capacity of an RNA species to undergo processing of its multimeric intermediates, either plus strands only or both the plus and minus strands. As shown by Sheldon and Symons (1993), the bypassing of the self-cleavage of the minus strands by vLTSV (lucerne transient streak virus satellite RNA) is not lethal, and hence might be an easy step in the evolution of a species. Therefore, the future identification of viroids localized in the chloroplast with asymmetric replication is possible. As for satellite RNAs, their replication mode could reveal a relationship without being a feature shared by all chloroplastic viroids. The cellular localization of viroid replication appears as a fundamental difference between viroid groups, and may have led to the evolution of two highly different clusters of viroids (i.e. group A and B). Consequently, we propose cellular localization as an essential criteria for group classification; chloroplast for group A and the nucleus for group B. Group A viroids can be further split into sub-groups based on the accumulation of their replication intermediates (i.e ASBVd accumulates multimeric strands while PLMVd and CChMVd do not).

Materials and methods

RNA isolation

RNA samples were isolated from leaves harvested from a total of 9 peach cultivars, 6 that had been infected by PLMVd and 3 healthy ones (see Table II). Four different procedures were used to prepare the RNA samples. i. The first procedure is a modified version of the procedure of Pallas *et al.* (1987) involving a polyethylene glycol precipitation (PEG-precipitation). Frozen leaf petioles (0.6 g) were ground to a fine powder in liquid nitrogen. The resulting powdered tissue was transfered to a 50 mL centrifuge tube and 5 ml phenol mix (phenol saturated Tris pH 8.0/ chloroform/ octanol; 100:100:4), 4 mL LG buffer (3.5 M LiCl and 0.3 M glycine pH 9.5), 80 µL bentonite solution, 40 µL 2-mercaptoethanol and 40 µL 20% LDS (lithium dodecyl sulphate in water) added. The mixture was vortexed for 30-60 sec, and then centrifuged at 10 000 g for 20 min. The resulting supernatant was transfered to a fresh tube and powdered PEG-6000 added to a final concentration of 20% (w/v). After vortexing for 15-30 sec, the tubes were successively incubated at 37°C for 5 min, room temperature for 10 min and on ice for 20 min prior vortexing again for 15 sec. The mixture was centrifuged at 10 000 g for 20 min, the pellet recovered, washed with 70% ethanol and air dried. The pellet was dissolved in 400 µL nuclease free water, transfered to a sterile eppendorf tube and centrifuged at 12 000 g for 5 min to remove any insoluble material. Finally, the RNA was ethanol precipitated. ii. The second procedure used the RNeasy Plant mini kit (Qiagen, Mississauga, Ont. Canada) to prepare RNA from ~150 mg of tissue according to the manufacturer's recommended protocol. iii. The third procedure was identical to that described in section ii. except that 5 mM EDTA was added to all buffers. iv. The fourth procedure was the Tris-EDTA extraction method of Hadidi *et al.* (1997). All RNA samples were quantified by UV spectrophotometry and electrophoresed on 1.3% agarose gels in order to assess the quality of the preparation. Dried RNA samples were stored at -70°C.

Preparation of RNA probes

The synthesis and purification of both the plus and minus strand-specific riboprobes (or antisense and sense probes, respectively) were performed using plasmid pPD1 as template (Beaudry *et al.*, 1995). Briefly, this construction possesses two tandemly repeated PLMVd sequences cloned into the *Pst* I restriction site of pBluescript II KS. The insert is flanked by T3 and T7 promoters for the production of plus and minus polarity transcripts, respectively. For Northen blot hybridization, we used the StripEZ transcription kit (Ambion Inc., Austin, TX) in order to obtain probes that can be stripped under mild washing conditions so as to permit multiple probings of the same membrane. The transcription reaction was performed in the presence of 50 µCi of [α-^{32}P]UTP (or GTP) (3000 Ci/ mmol, Amersham Life Science, Arlington Heights, IL). For *in situ* hybridizations, *in vitro* transcription was performed in the presence of 3.5 mM digoxigenin-11-UTP (Boehringer Mannheim, Laval, Que, Canada) according to the manufacturer's recommended conditions. During transcription, RNA of both polarities possessing hammerhead sequences are produced and self-cleave efficiently producing 338 nucleotide (nt) linear monomeric transcripts. After transcription, DNase (RNase free; Pharmacia, Uppsala, Sweden) treatment was performed and the nucleic acids ethanol precipitated. In general the transcripts were resuspended in 20 µL water and 10 µL of stop buffer (0.3% (w/v) each of bromophenol blue and xylene cyanol, 10 mM EDTA (pH 7.5), 97.5% (v/v) deionized formamide) added, and the resulting mixture denatured for two min at 65°C prior to electrophoresis through a 5% (w/v) polyacrylamide gel (PAGE) in 100 mM Tris-borate (pH 8.3), 1 mM EDTA, 7M urea buffer. Transcripts were detected by UV shadowing, excised, eluted, precipitated, purified by passage through Sephadex G-50 spun columns (Pharmacia, Uppsala, Sweden), lyophilized, quantitated by absorbance spectrophotometry at 260 nM and stored dry at -70°C. Incorporation of the DIG-label into the transcripts was monitored using dot blots probed with alkaline phosphatase-conjugated anti-DIG antibody (Boehringer Mannheim)

according to the manufacturer's procedure for DIG labelled transcripts, while Cherenkov counting was used for the ^{32}P-labelled transcripts.

PSTVd probes labeled with DIG were prepared by *in vitro* transcription from plasmid pHa106 (kindly provided by Dr. Martin Tabler; Tabler and Sänger, 1985). *Eco* RI-linearized pHa106 was transcribed with SP6 RNA polymerase producing a plus polarity 406 nt PSTVd riboprobe which is slightly longer than PSTVd monomer (i.e. 359 nt).

In situ hybridization for electron and light microscopy

Fixation and embedding of peach samples was performed as described by Lima *et al.* (1994) with minimal modifications. Briefly, leaf pieces from both healthy and PLMVd-infected peach plants (Siberian C cultivar) were fixed using a modified Karnovsky fixation (1% glutaraldehyde, 2% paraformaldehyde in 100 mM cacodylate buffer pH 7.2 containing 5 mM $CaCl_2$) for 2-48 hr, dehydrated through an ascending ethanol series at room temperature and embedded in LR Gold (Electron Microscopy Sci., Ottawa, Ont., Canada) at -20°C under UV light.

In situ hybridizations were performed as described by Egger *et al.* (1994). For light microsocpy, 0.5 µm sections were collected on poly-L-lysine-coated superfrost glass slides (dried 3 hr at 42°C) and acetylated. Prehybridization (1 hr) and hybridization (overnight) were performed as described for Northern blots (see below) except that the salmon sperm DNA was replaced by 0.5 mg/mL yeast tRNA, the temperature was held constant at 55°C, and both steps were performed in a petri dish humidified with 50% formamide/ 4 X SSC (20 X SSC is 3 M NaCl/ 0.3 M sodium citrate, pH 7.0). After hybridization, the slides were washed three times in 2X SSC at 37°C for 15 min, and once in 1X SSC for 15 min at 37°C. Identical results were obtained when the slides were washed after the first 2X SSC wash with 2X SSC containing 1 µg/mL RNase for 10 min at room temperature. Detection of DIG-labeled hybrids was performed as recommended by the manunfacturer using the reaction between alkaline phosphatase-coupled anti-digoxigenin antibody and indoxyl-

nitroblue tetrazolium in the presence of polyvinyl alcohol (DeBlock and Debrouwer, 1996). For electron microscopy, ultra-thin sections were mounted on 150 mesh nickel grids covered with parlodion-carbon-coated gold grids, and were then acetylated and washed twice with 2X SSC for 5 min. The grids were placed, section down, on 200-400 µL of hybridization buffer and incubated overnight at 60°C as described for light microscopy. After hybridization, the sections were washed at 37°C successively once for 20 min, and 4 times for 5 min with 2X SSC solution, and finally twice for 5 min with phosphate buffer saline (PBS: 0.34 M NaCl/ 0.007 M KCl/ 0.019 M KH_2PO_4/ 0.004 M Na_2HPO_4) with 0.1% tween-20. For indirect immunodetection, the sections were incubated for 60 min at 37°C with anti-DIG gold-labeled antibodies (from mouse; Boehringer Mannheim) diluted 1:25 in PBS/ 0.1% tween-20/ 1% casein. After two washes in PBS buffer, section were incubated for 60 min at 37°C on gold-labeled mouse anti-DIG (10 nm gold-labeled; Cedar lane, Toronto, Ont., Canada), and then briefly rinsed with distilled water. Grids were examined in a Phillips 300 electron microscope (Holland).

Northern and dot blot hybridizations

Northern blot hybridizations were performed as described previously (Fourney *et al.*, 1988) using probes radiolabeled with [α-^{32}P]UTP. Isolated RNA (2 µg) or PLMVd transcripts (0.5 ng) were either resuspended in formaldehyde gel-running buffer, denatured for 3 min at 80°C and fractionated on 1.3 % agarose gels containing formaldehyde and ethidium bromide (for visualization), or were resuspended in formamide/dye mixture (95% formamide/ 10 mM EDTA/ 0.05% bromophenol blue and 0.05% xylene cyanol), denatured for 2 min at 70°C and fractionated on 5% polyacrylamide gels (PAGE) using 50 mM Tris-borate (pH 8.3)/ 1 mM EDTA/ 7 M urea running buffer as previously described (Beaudry *et al.*, 1995). Nucleic acids were transfered by capillarity overnight to nylon filters (Hybond N+; Amersham Life Science) in either 10X (acrylamide gels) or 20X SSC (agarose gels). The filters were then UV crosslinked for 5 min and baked at 80°C for 90 min prior to being

prehybridized for 4 hr at 60°C in a solution containing 50% formamide/ 5X SSC/ 1% sodium dodecyl sulphate (SDS)/ 5% Denhardts/ 100 μg/mL salmon sperm DNA. Hybridizations were performed overnight (~16 hours) at 60°C with fresh prehybridization buffer containing the probe. After hybridization, the filters were successively washed 3 X 15 min in 1 X SSC/ 0.1 % SDS at 65°C, and 1 X 15 min in 0.1 X SSC/ 0.1 % SDS at 60°C (or 65°C). The filters were analyzed by either autoradiography or with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). All Northern blot hybridizations were performed as described above with the exception of a few controls which are indicated in the Results section. All blots were successively hybridized with the plus and the minus strand probes. Linear PLMVd transcripts used as controls were synthesized as described previously (Beaudry *et al.*, 1995). Circular PLMVd transcripts having only 3',5'-phosphodiester bonds were synthesized as reported (Beaudry and Perreault, 1995).

Dot blot hybridizations were performed as described for Northern blots with the exception that serial dilutions of the RNA samples were applied to the filter under vacuum. In addition, quantities ranging from 0.01 ng to 32 ng of gel-purified monomeric PLMVd transcripts of both polarities (synthesized by run-off transcription) were used as standards. Gel-purified monomeric (338 nt) PLMVd transcripts of both polarities were used as probes, and the filters quantified using a PhosphorImager (Molcular Dynamics).

Acknowledgements

This work was sponsored by a grant from Natural Sciences and Engineering Research Council (NSERC) of Canada to J.P.P. F.B. was the recipient of an NSERC studentship. J.P.P. is a Medical Research Council (MRC, Canada) scholar.

References

Beaudry,D. and Perreault, J.P. (1995) An efficient strategy for the synthesis of circular RNA molecules. *Nucleic Acids Res.*, **23**, 3064-3066.

Beaudry,D., Bussière, F., Lareau,F., Lessard,C. and Perreault,J.-P. (1995) The RNA of both polarities of the peach latent mosaic viroid self-cleaves *in vitro* solely by single hammerhead structures. *Nucleic Acids Res.*, **23**, 745-752.

Bonfiglioli,R.G., Mcfadden,G.I. and Symons,R.H. (1994) *In situ* hybridization localizes avocado sunblotch viroid on chloroplast thylakoid membranes and coconut cadang cadang viroid in the nucleus. *Plant J.*, **6**, 99-103.

Bonfiglioli,R.G., Webb,D.R. and Symons,R.H. (1996) Tissue and intra-cellular distribution of coconut cadang cadang viroid and citrus exocortis viroid determined by *in situ* hybridization and confocal laser scanning and transmission electron microscopy, *Plant J.*, **9**, 457-465.

Branch,A.D., Benenfeld,B.J. and Robertson,H.D. (1988) Evidence for a single rolling circle in the replication of potato spindle tuber viroid. *Proc. Natl. Acad. Sci. USA*, **85**, 9128-9132.

Bussière,F., Lafontaine,D., Côté,F., Beaudry,D. and Perreault,J.-P. (1995) Evidence for a model ancestral viroid. *Nucl. Acids. Symp. Ser.*, **33**, 143-144.

Bussière,F., Lafontaine,D. and Perreault,J.-P. (1995) Conmpilation and analysis of viroid and viroid-like RNA sequences. *Nucleic Acids Res.*, **24**, 1793-1798.

Colpan,M., Schumacher,J., Brüggemann,W., Sänger,H.L., Reisner,D. (1983) Large-scale purification of viroid RNA using Cs_2SO_4 gradient centrifugation and high-performance liquid chromatography. *Anal. Biochem.*, **131**, 257-265.

Côté,F. and Perreault,J.-P. (1997) Peach latent mosaic viroid is locked by a 2',5'-phosphodiester bond produced by *in vitro* self-ligation. *J. Mol. Biol.*, **273**, 533-543.

Daros,J.-A., Marcos,J.F., Hernandez,C. and Flores,R. (1994) Replication of avocado sunblotch viroid: Evidence for a symmetric pathway with two rolling circles and hammerhead ribozyme processing. *Proc. Natl. Acad. Sci. USA*, **91**, 12813-12817.

Daros,J.-A. and Flores,R. (1995) Identification of a retroviroid-like element from plants. *Proc. Natl. Acad. Sci. USA*, **92**, 6856-6860.

Davies,C., Sheldon,C.C. and Symons,R.H. (1991) Alternative hammerhead structures in the self-cleaving of avocado sunblotch viroid RNAs. *Nucleic Acids Res.*, **19**, 1893-1898.

DeBlock,M. and Debrouwer, D. (1996) RNA-RNA *in situ* hybridization using DIG-labeled probes: the effect of high molecular weight polyvinyl alcohol on the alkaline phosphatase indoxyl-nitroblue tetrazolium reaction. In *Nonradioactive in situ hybridization application manuel*, 2nd ed. Boehringer Mannheim, p.141-145.

Di Serio,F., Daor,J.A., Ragozzino,A. and Flores,R. (1997) A 451-nucleotide circular RNA from cherry with hammerhead ribozymes in its strands of both polarities. *J. Virol.* **79**, 6603-6610.

Egger,D., Troxler,M. and Bienz,K. (1994) Light and electron microscopic *in situ* hybridization: non-radioactive labeling and detection, double hybridization, and combined hybridization-immunocytochemistry. *J. Histochem. Cytochem.*, **42**, 815-822.

Elena,S.F., Dopazo,J., Flores,R., Diener,T.O. and Moya,A. (1991) Phylogeny of viroids, viroidlike satellite RNAs, and the viroidlike domain of hepatitis d virus RNA. *Proc. Natl. Acad. Sci. USA*, **88**, 5631-5634.

Fourney,R.M., Miyakoshi,J., Day III,R.S. and Paterson,M.C. (1988) Northern blotting: efficient RNA staining and transferet. *Focus*, **10**, 5-6.

Flores,R., Di Serio,F. and Hernandez,C. (1997) Viroids: The noncoding genomes. *Semin.Virol.*, **8**, 65-73.

Forster,A.C., Davies,C., Sheldon,C.C., Jeffries,A.C. and Symons,R.H. (1988) Self-cleaving viroid and newt RNAs may only be active as dimers. *Nature*, **334**, 265-267.

Hadidi,A., Giunchedi,L., Schamboul,A.M., Poggi-Pollini,C. and Amer,M.A. (1997) Occurrence of peach latent mosaic viroid in stone fruits and its transmission with contaminated blades. *Plant Dis.*, **81**, 154-158.

Harders,J., Lukacs,N., Robert-Nicoud,M., Jovin, J.M. and Riesner,D. (1989) Imaging of viroids in nuclei from tomato leaf tissue by *in situ* hybridization and confocal laser scanning microscopy. *EMBO J.*, **8**, 3941-3949.

Hernandez,C. and Flores,R. (1992) Plus and minus RNAs of peach latent mosaic viroid self-cleave *in vitro* via hammerhead structures. *Proc. Natl. Acad. Sci. USA*, **89**, 3711-3715.

Hutchins,C.J., Keese,P., Visvader,J.E., Rathjen,P.D., McInnes,J.L. and Symons,R.H. (1985) Comparison of multimeric plus and minus in chloroplasts of avocado leaves by *in situ* hybridization. *Arch. Virol.*, **138**, 385-390.

Lafontaine,D., Beaudry,D., Marquis,P. and Perreault,J.-P. (1995) Intra- and intermolecular nonenzymatic ligations occur within transcripts derived from the peach latent mosaic viroid. *Virology*, **212**, 705-709.

Lareau,F. (1996) Étude d'éléments de la réplication du viroïde de la mosaique latente du pêcher. in *Master thesis*, Université de Sherbrooke.

Lima,M.I., Fonseca,M.E.N., Flores,R. and Kitajima,E.W. (1994) Detection of avocado sunblotch viroid in chloroplasts of avocado leaves by *in situ* hybridization. *Arch. Virol.*, **138**, 385-390.

Link,G. (1996) Green life: control of chloroplast gene transcription. *Bioessay*, **18**, 465-471.

Liu,Y.-H. and Symons,R.H. (1998) Specific RNA self-cleavage in coconut cadang cadang viroid: potential for a role in rolling circle replication. *RNA*, **4**, 418-429.

Marcos,J.F. and Flores, R. (1992) Characterization of RNAs specific to avocado sunblotch viroid synthesized *in vitro* by a cell-free system from infected avocado leaves. *Virology*, **186**, 481-488.

Navarro,B. and Flores,R. (1997) Chrysanthenum chlorotic mottle viroid: unusual structural properties of a subgroup of self-cleaving viroids with hammerhead ribozymes. *Proc. Natl. Acad. Sci. USA*, **94**, 11262-11267.

Pallas,V., Navarro,A. and Flores,R. (1987) Isolation of a viroid-like RNA from hop different from hop stunt viroid. *J. Gen. Virol.*, **68**, 3201-3205.

Schumacher,J., Sanger,H.L. and Riesner,D. (1983) Subcellular localization of viroids in highly purified nuclei from tomato leaf tissue. *EMBO J.*, **2**, 1549-1555.

Sheldon,C.C. and Symons,R.H. (1993) Is hammerhead self-cleavage involved in the replication of a virusoid *in vivo*? *Virology*, **194**, 463-474.

Silver,S.L., Rasochova,L., Dineshkumar,S.P. and Miller,W.A. (1994) Replication of barley yellow dwarf virus satellite RNA transcripts in oat protoplasts. *Virology*, **198**, 331-335.

Symons, R.H. (1997) Plant pathogenic RNAs and RNA catalysis. *Nucleic Acids Res.*, **25**, 2683-2689.

Tabler,M. and Sänger,H.L. (1985) Infectivity studies on different potato spindle tuber viroid (PSTV) RNAs synthesized *in vitro* with the SP6 transcription system. *EMBO J.*, **4**, 2191-2199.

Description du travail

Les objectifs principaux de l'article étaient d'identifier la localisation cellulaire de PLMVd et de caractériser les différentes formes du viroïdes retrouvées à l'intérieur des cellules infectées. Le patron d'accumulation nous permet ainsi de connaître le type de cercle roulant utilisé par PLMVd *in vivo*. Les résulats montrent que PLMVd, tout comme ASBVd, est un viroïde chloroplastique qui se réplique par un mécanisme en cercle roulant symmétrique. Dans les extraits d'ARN on a également identifié des ARN qui possèdent vraisemblablement de fortes similarités de séquences avec PLMVd.

Implication dans le travail

J'ai débuté les études d'hybridation *in situ*. J'ai par la suite supervisé le travail de J. Lehoux qui a poursuivi les études de localisation cellulaire. J'ai réalisé le reste des travaux présentés dans l'article en plus de participer activement à la rédaction du manuscrit.

CHAPITRE 3:

Études structurales de PLMVd

3.1 ARTICLE:

Mapping in solution shows the peach latent mosaic viroid to possess a new pseudoknot in a complex, branched secondary structure

Frédéric Bussière, Jonathan Ouellet, Fabien Côté,

Dominique Lévesque et Jean-Pierre Perreault

Département de Biochimie

Université de Sherbrooke

Sherbrooke, Québec, J1H 5N4, Canada

Article accepté dans: *Journal of Virology*

ABSTRACT

We have investigated the secondary structure of peach latent mosaic viroid (PLMVd) in solution, and present here the first description of the structure of a "branched" viroid in solution. Based on both the circularly permuted RNA method and the exploitation of RNA internal secondary structure to position 5' and 3' termini, different PLMVd transcripts of plus polarity were produced and studied by nuclease mapping and binding shift assays using DNA and RNA oligonucleotides. We show that PLMVd folds into a complex, branched secondary structure in which 86 (25.4 %) and 252 (74.5 %) nucleotides are in single- and double-stranded regions, respectively. More importantly, this structure includes a novel pseudoknot that is conserved in all PLMVd isolates, and appears as a potential structural motif of the chrysanthemum chlorotic mottle viroid (CChMVd). The presence of this pseudoknot is supported by nuclease mapping of a mutant PLMVd transcript, and by the base conservation observed in all natural PLMVd variants. This pseudoknot should allow PLMVd and CChMVd, which belong to the same viroid sub-group (i.e. PLMVd sub-group), to fold into very compact structures.

Keywords: viroid / RNA structure / RNA organism / catalytic RNA

INTRODUCTION

Viroids are small (~300 nucleotides), single-stranded, circular RNAs that infect higher plants causing significant losses in the agricultural industry (see refs. 1 and 2 for reviews). Viroids have been classified in two groups (i.e. A and B) based primarily on whether or not they possess five typical structural domains found in the group B viroids (1). Further division among the group B members depends on the sequence and length of the conserved central region (CCR). Viroids that do not possess any kind of sequence and structural similarity with the group B viroids have been classified as belonging to group A. The viroids from this group possess self-cleaving hammerhead motifs that are crucial for their replication via a rolling circle mechanism.

The group A viroids include the avocado sunblotch viroid (ASBVd), the peach latent mosaic viroid (PLMVd) and the chrysanthemum chlorotic mottle viroid (CChMVd) (3). Both PLMVd and CChMVd have been proposed to adopt branched secondary structures (Fig. 1A) instead of the rod-like ones proposed for most viroids, including ASBVd (3). The unusual conformations of PLMVd and CChMVd are supported by their insolubility in 2 M lithium chloride, whereas ASBVd and a number of non self-cleaving viroids (i.e. the group B viroids) are soluble in this high salt solution (3). In general, secondary structures of viroids are predicted using computer software, and are useful for the formulation of hypotheses on the structure/function relationships of these RNA molecules (4). Characterization of biological structures *in vitro* as well as *in vivo* is obviously more accurate for elucidating the structure/function relationship. The only secondary structure of a viroid that has been extensively studied in solution is that of the potato spindle tuber viroid (PSTVd) (5). This group B species, which was shown to adopt a rod-like shape in solution, is responsible for most of our knowledge on the biology of viroids.

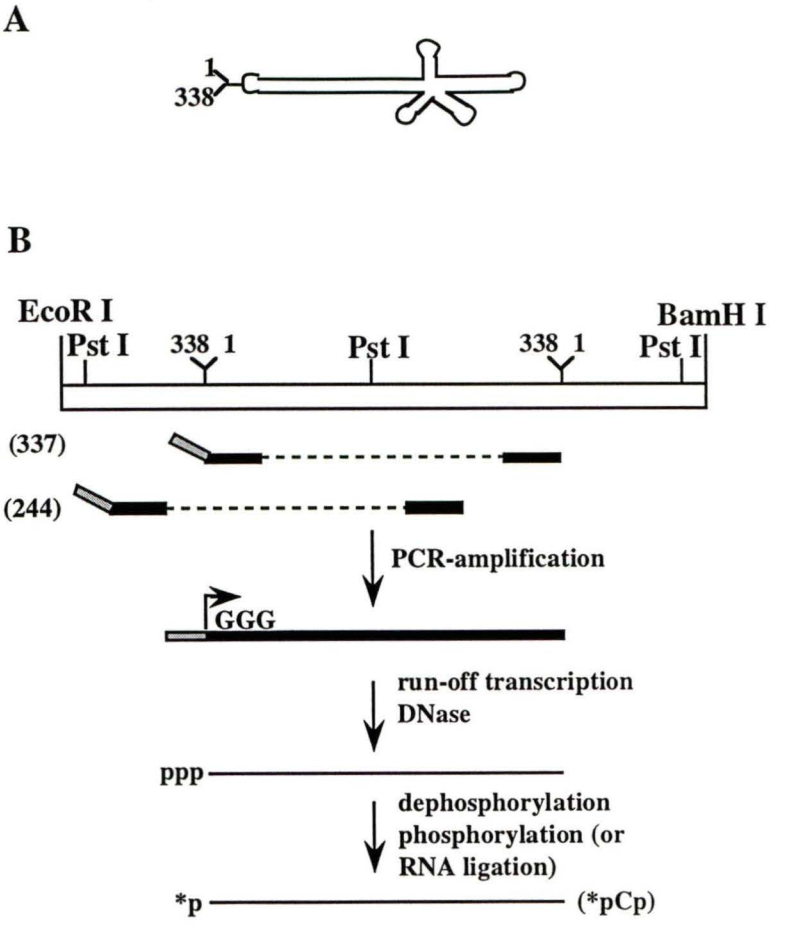


Figure 1. Strategy of synthesis of the PLMVd strands. A Schematic representation of the circular PLMVd strand of plus polarity. By convention position 1 was arbitrarily fixed in the left hand loop. **B** Strategy of synthesis of the linear PLMVd strands and their end labelling based on the circularly permuted RNA method. The 5'-end positions of the transcripts used in this work are indicated in parentheses.

In order to determine the secondary structure of PLMVd (the RNA species that causes peach latent mosaic disease (6)) in solution, we performed nuclease mapping and oligonucleotide binding shift assays on PLMVd synthesized by *in vitro* transcription. The results confirm that PLMVd folds into a complex, branched secondary structure, and show that it can form a novel pseudoknot.

MATERIALS AND METHODS

Synthesis of DNA templates. DNA templates for *in vitro* transcription of the three different RNA molecules used in this study were synthesized from the pPD1 vector (7). Briefly, this construct possesses two tandemly repeated PLMVd sequences (from the Armking peach cultivar, ref. 6) cloned into the *Pst* I restriction site of pBluescript II KS. The strategy for the production of monomeric PLMVd RNA has been reported previously (8). DNA templates used for *in vitro* transcription were amplified from pPD1 with Pwo DNA polymerase (Boehringer Mannheim) using either forward (5'-TAATACGACTCACTATAGGGTCAAAAGTTTCGCCGC-3') and reverse (5'-TATGAGTTTCGTCTCATTTC-3') primers for RNA transcripts starting from position 337, or forward (5'-TAATACGACTCACTATAGGGATTCAAACCCGGTC-3') and reverse (5'-GGGTAGACGTCGTAATCC-3') primers for RNA transcripts starting from position 244. The DNA template used for the production of the mutant harboring the sequence $_{212}$AAAA$_{215}$ was initially prepared from pPD1 using the PCR product generated with the forward primer used for transcripts starting at position 337 with a mutated reverse primer (5'-TTTCTACG<u>TTTT</u>TACCTGGA-3'; the underlined nucleotides are the mutant ones); and, secondly, using the PCR product generated with a mutated forward primer (5'-TCCAGGTA<u>AAAA</u>CGTAGAAA-3') coupled with the reverse primer used for the transcripts starting at position 337. These two PCR products were mixed together and amplified using Taq DNA polymerase in order to produce full sized DNA template, which

was then verified by dideoxynucleotide chain termination sequencing using the T7 sequencing kit (Amersham Pharmacia Biotech).

Run-off transcription. *In vitro* transcription reactions contained unpurified PCR products, 35 units RNAguard RNase inhibitor (Amersham Pharmacia Biotech), 4 mM of each ribonucleotide, 80 mM HEPES-KOH, pH 7.5, 24 mM $MgCl_2$, 2 mM spermidine, 40 mM dithiothreitol, 0.02 unit of pyrophosphatase (Boehringer Mannheim) and 25 μg of purified T7 RNA polymerase (9,10) in a final volume of 100 μL, and were incubated at 37°C for 4 h. After transcription, 10 units of RNase-free DNase RQ1 (Promega) were added and the reaction incubated at 37°C for 15 min in order to remove the template DNA. The reaction mixtures were then extracted with phenol:chloroform (1 vol: 1 vol), and the nucleic acids ethanol precipitated, ethanol washed, and fractionated by denaturing 5% polyacrylamide gel electrophoresis (19:1 ratio of acrylamide to bisacrylamide) in buffer containing 90 mM Tris borate, pH 7.5, 7 M urea, and 2 mM EDTA. The reaction products were visualized by UV shadowing, and the bands corresponding to the correct size (338 nt) cut out. The transcripts were then eluted from the gel slices by incubation in 0.1% sodium dodecyl sulfate and 0.5 M ammonium acetate solution overnight at 4°C. The resulting solution was then passed through a G-50 Sephadex spun column (Amersham Pharmacia Biotech), and precipitated by the addition of 0.1 volume of 3M sodium acetate, pH 5.2 and 2.0 volumes of ethanol. Transcript yield was determined by spectrophotometry at 260 nm.

5'- and 3'-end labelling of transcripts. Transcripts (4 μg) were 5'-end labelled by incubating at 37°C for 30 min in a final volume of 100 μL containing 250 mM Tris-HCl, pH 8.3, 35 units RNAguard, and 0.2 units calf intestinal alkaline phosphatase (Boehringer Mannheim). After phenol extraction and ethanol precipitation, the pellet was resuspended in 10 μL H_2O and incubated for 60 min at 37°C in a final volume of 50 μL containing 1x One-Phor-All buffer (10 mM Tris-acetate pH 7.5, 10 mM magnesium acetate and 50 mM potassium acetate, Amersham Pharmacia Biotech), 35 units RNAguard, 40 μCi

[γ-^{32}P]ATP (3000 mCi/mmol, Amersham), and 30 units T4 polynucleotide kinase (Amersham Pharmacia Biotech).

Transcripts (4 µg) were 3'-end labelled by incubating in a final volume of 20 µL containing T4 RNA ligase buffer (50 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$, 10 mM dithiothreitol and 1 mM ATP; New England Biolabs), 10% dimethylsulfoxide, 0.02% bovine serum albumine, 40 µCi [^{32}P]pCp (3000 mCi/mmol; Amersham), 35 units RNAguard and 40 units T4 RNA ligase (New England Biolabs) at 10°C overnight. The reactions were quenched by the addition of 12 µL of stop solution (97% deionized formamide, 10 mM EDTA pH 7.5, 0.2 % xylene cyanol), and the mixtures heat denatured and separated on denaturing 5% PAGE gels. Full length transcripts (338 nt +/-pCp) were detected by autoradiography, excised, eluted, and passed through a G-50 Sephadex spun colunm as described above. The concentration of RNA was determined by Cherenkov counting.

Nuclease mapping. Either 5'- or 3'-end labelled PLMVd transcripts (~50 000 cpm) were resuspended in a 4 µL of either 10 mM Tris-HCl pH 8.0 / 1 mM $MgCl_2$ (low salt) or 20 mM Tris-HCl pH 8.0 / 10 mM $MgCl_2$ / 100 mM NH_4Cl (high salt) solution. The transcripts were then denatured by heating at 65°C for 2 min, and renatured by slow cooling to room temperature. This procedure of denaturation-renaturation was performed prior to all enzymatic reactions unless indicated otherwise in the text. Either RNase T_1 (0.25 units; Amersham Pharmacia Biotech), RNase T_2 (6 units; Gibco BRL), RNase Phy M (0.5 units; Pharmacia), or RNase V_1 (1 unit; Amersham Pharmacia Biotech) were then added (in a final volume of 6 µL), and the mixtures incubated at either 25°C or 37°C for various times. For the RNase U_2 reactions (0.02 units; Amersham Pharmacia Biotech), the transcripts were resuspended in a buffer containing 20 mM sodium citrate pH 6.0. These mixtures were incubated at 37°C, and aliquots removed at various times and quenched by the addition of 5 µL stop solution. Ribonuclease T_1 digestion was also performed in the presence of 7M urea (at 55°C). Alkaline hydrolysis was performed so as to permit accurate

assignment of the cleavage sites. The reaction products were fractionated on either 5%, 8% or 12% denaturing PAGE gels, and then either exposed to a X-ray film, or fixed, dried and exposed to a PhosphorImager screen when quantification was required.

Oligonucleotide binding shift assays. RNA oligonucleotides were chemically synthesized using an automated oligonucleotide synthesizer (Keck Biotechnology Resource Laboratory, Yale University), deprotected and gel-purified as described previously (11). Deprotected and purified DNA oligonucleotides were purchased from Gibco-BRL. Each oligonucleotide (5 pmol) was 5'-end labelled using T4 polynucleotide kinase in the presence of an excess of [γ-^{32}P]ATP as described above, and was purified by phenol-chloroform extraction followed by passage throw two Sephadex G-50 spun columns. The quantity of labelled oligonucleotide was determined by Cherenkov (cpm) counting, while the quality was verified by denaturing 20% PAGE. Dried PLMVd transcripts (500 ng) were resuspended in 8 µL of a solution containing 10 mM Tris-HCl pH 8.0 / 1 mM $MgCl_2$ (low salt). Radiolabelled oligonucleotide (20 000 cpm; 2 µL) was added either before or after the transcripts were denatured for 2 min at 65°C, and slowly cooled to 4°C to renature them. The mixtures were then incubated at 4°C for at least 60 min prior to the addition of 2 µL of agarose-dye (30% glycerol [vol:vol], 1 mM EDTA pH 8.0, 0.25% xylene cyanol and bromophenol blue) and analysis on 2.5% agarose gels. The resulting gels were fixed, dried and exposed to X-ray films.

RESULTS

RNA synthesis strategy. In order to probe the secondary structure of a PLMVd sequence variant isolated from the peach Armking cultivar (also named french variant; ref. 6), a unique ^{32}P-labelled phosphate has to be introduced at an end of the RNA strand. Initially, we tried to label PLMVd monomeric strands produced by the self-cleavage of linear concatamers. However, these RNA strands were inefficiently ^{32}P-labelled at both the 5'- and 3'-ends, suggesting that both extremities were located inside the molecule. Consequently, we developped a second strategy for the production of full-length PLMVd molecules that can be efficiently labelled at both ends (8). This strategy is based on the circularly permuted RNA method, and on the exploitation of RNA internal secondary structure to position the 5' and 3' termini such that they will be located in single-stranded regions which are easily accessible for efficient end-labelling (Fig. 1B). In this approach, non-radioactive, full length PLMVd transcripts possessing a 5'-triphosphate (at position 337) and 3'-hydroxyl (at position 336) were produced and further purified. Note that positions 336 and 337 are located in the left-end loop (Fig. 1A). The only modification as compared to the wildtype sequence is the introduction of three guanosines at the 5'-end in place of the $_{337}$UCA$_{1}$ so as to allow for efficient transcription. A second PLMVd-derived RNA molecule whose 5'-end is located at position 244 was produced with the appropriate oligonucleotides according to the same procedure. Since positions 244-246 correspond to three guanosines in the wildtype sequence, no mutation was required in order to favor *in vitro* transcription. This transcript will confirm the results from the previous construct, and more specifically will allow study of the secondary structure of the left hand loop, including positions 336-337 where the other construct was opened.

Nuclease probing. The structure of the plus polarity PLMVd strands were probed using various nucleases which cleave 3' of different nucleotides: i. ribonuclease (RNase) T_1, which preferentially cleaves single-stranded guanosines; ii. RNase U_2, which

preferentially cleaves single-stranded adenosines; iii. RNase Phy M, which preferentially cleaves single-stranded adenosines and uracils; iv. RNase T_2, which preferentially cleaves single-stranded nucleotides regardless of base identity; and, v. RNase V_1, which preferentially cleaves double-stranded nucleotides regardless of base identity. In addition, some portions of the molecules were probed with the *Bacillus cereus* RNase, which has a preference for single-stranded cytosines and uracils, in order to confirm the data obtained with the other enzymes. The reaction conditions for each nuclease were optimized prior to the structural probing. This optimization was required as no carrier (i.e. tRNA) was added to the reactions so as to avoid any structural interference with the PLMVd strands. Regardless of whether the RNase mapping was performed at 22°C or 37°C, no significant differences were detected in the cleavage pattern.

Prior to the nuclease probing, the transcripts were treated in several ways in order to favor structural homogeneity. For example, they were denatured at 65°C and then renatured by slow cooling to either room temperature or 4°C; or, they were denatured at 85°C (always in the absence of magnesium to avoid self-cleavage) and then renatured at room temperature in a two step procedure (i.e. 2 min at 55°C followed by 2 min at room temperature) prior to addition of the magnesium. With the exception of the observation that incubation at 85°C stimulates self-cleavage of the transcripts, the various treatments all produced the same digestion pattern (data not shown). Therefore, we decided to use the simplest treatment consisting of a 2 min denaturation at 65°C followed by slow cooling to room temperature. During nuclease hydrolysis experiments aliquots were removed from the reaction at different times, stopped and fractionated on 5%-12% polyacrylamide sequencing gels. Figure 2 shows a typical autoradiogram for 5'-end-labelled PLMVd transcripts starting at position 337. Under these electrophoretic conditions the region from positions 29 to 66, which corresponds to a portion of the upper strand of the left arm and the first hairpin on the top (i.e. P11 and P1 stems, respectively, Fig. 3), was revealed. Nucleotides 29 to 66 are primarily hydrolyzed by RNase V_1, indicating that they are in a double-stranded region.

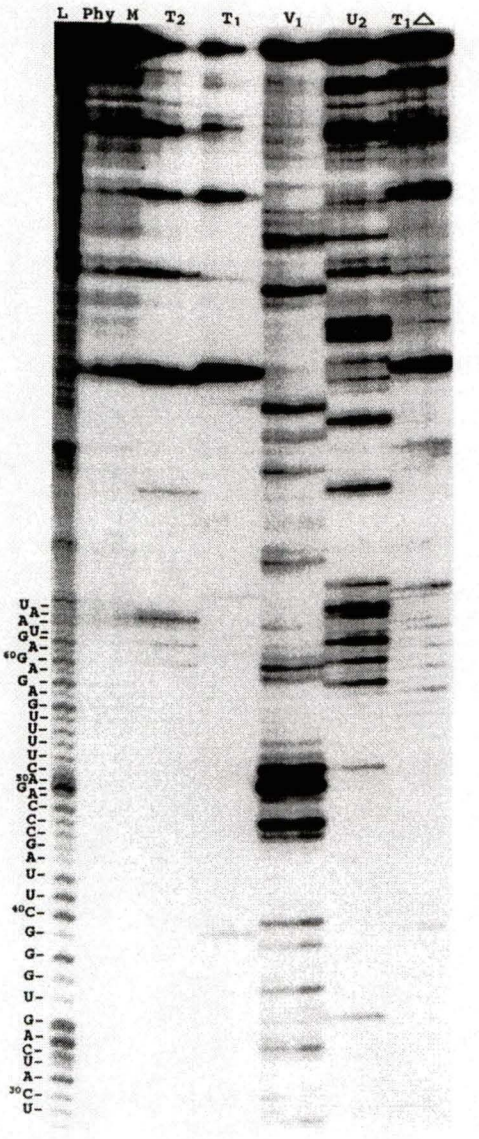


Figure 2. Autoradiogram of a PLMVd nuclease mapping assay analyzed on a 12% PAGE gel. The nucleases used are indicated at the top. L represents the alkaline hydrolysis ladder and T1Δ represents the RNase T1 reaction incubated at 55°C. Aliquots were removed at 30 sec, 1 min and 2 min (left to right) for each enzyme. The PLMVd sequence for the region from U29 to U66 is shown at the left.

TABLE 1. Compilation of band intensity observed by RNAse mapping[a]

nt	T_1	U_2	P.M.	T_2	V_1
1	A		3	4	3
2	U			4	3
3	C				3
4	A		3	3	4
5	A		3	3	4
6	A		3	3	3
7	A		2	1	2 1
8	G	1			3
9	U			1	3
10	U			1	4
11	U			1	4
12	C				4
13	G	1			1 3
14	C				3 2
15	C				2 2
16	G	2			3 1
17	C				1 2
18	A		1	1	2
19	U			1	3
20	U			1	3
21	U			2	1 3
22	C				1 3
23	A		1		1 3
24	G				4
25	C				2
26	G	1			1 2
27	G				3
28	C				3
29	U			1	1 2
30	C		1		2
31	A			1	2
32	U				1
33	C				2
34	A				3
35	G			1	2
36	U				2
37	G				1
38	G	1	1		2
39	G	2			1
40	C				4
41	U				2
42	U				2
43	A		1		1
44	G				2
45	C				3
46	C				1
47	C				3
48	A				2
49	G				4
50	A		1		3
51	C		1		3
52	U				2
53	U			1	2
54	U		1	1	1 1
55	U		1	1	1 1
56	G	1	1		2
57	A		1	1	2
58	G		1		2
59	A		1		3
60	G		1		1 3
61	A		3		1
62	G	1	1		1 2
63	U		1		1
64	A		3	1	1
65	A		3	1	3 1
66	U		2	1	1
67	G	2			1 1
68	A		3	1	1
69	C				2
70	C				2
71	U		1		2
72	C				2
73	U			1	3
74	C				1
75	A			1	3
76	G				2
77	C				2
78	C				2
79	C				2
80	C				2
81	U			1	1
82	C				2
83	C				3 1
84	A		3	1	1
85	C				2
86	C				2
87	U			1	1
88	U			1	1 1
89	G	1			1
90	G				2
91	G				3
92	G	1			2
93	U				2
94	G				2
95	C			1	2
96	C			1	2
97	C				1
98	U		1	1	1 1
99	A		2		1 1
100	U			1	1 1
101	U				2
102	C				3
103	G	1			2
104	G	1			2
105	A		1		2
106	G	1			2
107	C				2
108	A		2	1	3
109	C				4
110	U			3	4
111	G	4			4
112	C				4
113	A		2	2	2
114	G	2	1		2
115	U			1	2
116	U			1	2
117	C			1	1
118	C				3
119	C				3
120	G	1			2
121	A			1	2
122	U			1	1 2
123	A		2	1	1 2
124	G	1	1		1 2
125	A		1	2	2 1
126	A		3	1	2 1
127	A		2	1	2 1
128	G	1			1 1
129	G	1			2
130	C				1
131	U			2	1
132	A		2	2	1 1
133	A		1	2	
134	G	2	1		
135	A		2	1	1 1
136	A		1		2
137	C				4
138	C				4
139	U				3
140	C				2
141	G	3			2 2
142	C				3 2
143	A		4	4	4
144	A		4	4	3
145	U		1	4	3 1
146	G	1	2		2
147	A		1	2	3
148	G				2
149	G	1			2
150	U			1	2
151	A			2	1
152	A		1	2	1
153	G	1			1 2
154	G				2
155	U			1	2
156	G				2
157	G	2			1 2
158	G	2			2 1
159	A		2	3	1 2
160	C				1 2
161	U		1	2	2
162	U			2	2 1
163	U			1	1 2
164	U			2	1 2
165	C			2	2 2
166	C			2	2 2
167	U			2	2 2
168	U			1	2
169	C				1 1
170	G	4	2		4
171	G	3	1		2 1
172	G	1			1 2
173	A		1	1	2
174	A		1	1	2
175	C				3
176	C				2
177	A		1	1	2
178	A		3	2	2 1
179	G	1			1 2
180	C				2
181	G			1	2
182	G			1	3
183	U		1	2	2 1
184	U			2	1 2
185	G	1			1 2
186	G				3
187	U				3
188	U			1	2
189	C				2
190	C			2	1
191	G	2	1		1 2
192	A		1		2
193	G	1			2
194	G				3
195	G				3
196	G				2
197	G	1			1 2
198	G	2			2 2
199	U		1	2	1 2
200	G	1			2
201	U			1	3
202	G				3
203	A		2		3
204	U				4
205	C				4
206	C				4
207	A		1	1	3
208	G	1			1 2
209	G	2	1		2 1
210	U		1	2	2 1
211	A		2	2	2 1
212	C				2
213	C				3
214	G	2			1 3
215	C				2
216	C				1 1
217	G	4	1		3
218	U		3	4	3 1
219	A		1	4	2 1
220	G	4	1		3
221	A		3	3	3
222	A		3	2	3
223	A		2	2	4
224	C				4 1
225	U		1	1	4 2
226	G	4			1 2
227	G	2			3
228	A		1		2
229	U				3
230	U			1	2
231	A			1	2
232	C				1
233	G	3			1 1
234	A		2	1	3
235	C				3
236	G	2			2 1
237	U			1	1 2
238	C				2
239	U				2
240	A		1	1	2
241	C				3
242	C				2
243	C			1	
244	G	1			1 1
245	G	1			2
246	G				3
247	A		3		3
248	U			3	2
249	U			1	4
250	C			2	1
251	A		3	3	4
252	A		3	4	4
253	A		3	3	4
254	C				4
255	C				3
256	C			2	1
257	G	1	1		1 2
258	G				2
259	U			1	2
260	C				2
261	C				2
262	C				2
263	C			1	1
264	U			1	2
265	C				3
266	C				4
267	A		3	2	2
268	G	2			2
269	A		1		2
270	A		3	2	1 1
271	G	4	1		3 2
272	U			4	3 1
273	G	4	2		3 1
274	A		4	2	3 2
275	U		3	1	3
276	U				3
277	C				2
278	U				2
279	G	1			2
280	G	1			2
281	A			1	1 2
282	A		3	1	2 1
283	G	2			1 1
284	A		3	2	2 1
285	U			2	1 1
286	G	3			1 2
287	A		1		1 2
288	G	1			2
289	U			1	2
290	C				3
291	U				2
292	G	2			1 1
293	U		1	2	2 1
294	G	1	1		2
295	C				3
296	U				3
297	A			1	3
298	A			1	2
299	G	1			2
300	C			1	1
301	A		2	2	2
302	C				2
303	A				2
304	C				3
305	U				2
306	G				3
307	A		1		3
308	U				3
309	G				2
310	A				2
311	G				2
312	U		1		2
313	C			1	1
314	U			2	2 1
315	C				2
316	U			1	2
317	G	1			2
318	A				1 2
319	A				2
320	A			1	2
321	U				2
322	G	1			1 1
323	A		2	1	2
324	G	2			1 1
325	A		2	2	2
326	C				1
327	G	1			1 2
328	A		1	2	1 2
329	A			2	2 2
330	A			2	1 2
331	C				2 1
332	U		1	2	2 1
333	C				3 1
334	A		3	3	4 1
335	U		1	4	4 1
336	A		4	4	4
337	U			4	4
338	C				4 1

[a] Relative intensities were determined for each pattern of RNase hydrolysis, with 4 being the stronger. Rectangles indicate the probability of a nucleotide being either in a single- (open rectangle) or a double- (closed) stranded regions. Intermediate intensity (between closed and opened rectangles) shows a tendency for one or the other structure.

At least the first 180 nucleotides from both the 5'-^{32}P- and 3'-^{32}pCp-labelled transcripts were resolved using different migration conditions (e.g. by varying the migration times and acrylamide concentrations). The relative intensity of all radioactive bands was scored in comparison to the most intense bands found in each reaction. Scores varying from 1 to 4 (with 4 being the most intense) were attributed to each nucleotide, and then a relative intensity average at each nucleotide for each nuclease was calculated (Table I). In some cases, more than eight experiments were performed in order to ensure the reproducibility of the results. The two RNA constructs gave similar results with the exception that nuclease accessibility was greater close to their respective ends. For example, the construct starting at position 337 indicated that both ends were single-stranded and highly accessible; while the construct starting at 244 confirmed the presence of the left-hand loop, although the intensity of the bands was less.

Description of the proposed secondary structure. In order to facilitate the description of the secondary structure (Fig. 3) we have numbered the helices as if this circular RNA had been synthesized from position 1, which by convention is located in the left hand loop, and numbered the stems in the order of appearance. If two helices have the potential to stack, but are separated by internal loop(s), the same number was given to both stems and they are then differentiated by a letter (e.g. P11a-e stems, see Fig. 3).

According to this secondary structure, 86 (25.4 %) and 252 (74.5 %) nucleotides are located in single- and double-stranded regions, respectively, when GU wobble basepairs and Watson-Crick basepairs proven to exist in high salt conditions are considered (see below). This structure includes 11 helices (P1-P11), with the P11 stem including four consecutives stems (P11a to P11d) that form a left-hand rod-like domain that contains the hammerhead sequences of both the plus and minus polarities. While most of the helices correspond to hairpins (i.e. stem-loop structures), the P5 stem is an internal helix which acts as a bridge closing the molecule. The proposed structure includes 17 basepairs

illustrated with dotted lines indicating that it is not unequivocal whether or not the two nucleotides in question are single stranded or basepaired in the presence of low salt (50 mM Tris-HCl, 10 mM $MgCl_2$). In fact, both forms may co-exist. In order to verify this hypothesis, and to investigate potential structural transitions caused by higher ionic strengths, nuclease probing experiments were also performed in buffer containing 50 mM Tris-HCl, 10 mM $MgCl_2$ and 100 mM NH_4Cl (raw data not shown). In general, the results were identical to those reported in Table 1, regardless of the difference in the processing of each enzyme. No structural transition was detected. The uncertain basepairs (i.e. dotted lines in Fig. 3) in the middle of the P3 stem, in the bottom of the P5 stem, in the P6a stem and in the P10 stem appeared to be closed, eliminating the possibility of the co-existence of single- and double-stranded structures in higher salt conditions. However, the uncertain basepairs could also be due to the steric hindrance of RNases because of their large size. Thus, the proposed secondary structure appears to remain identical under both salt conditions tested.

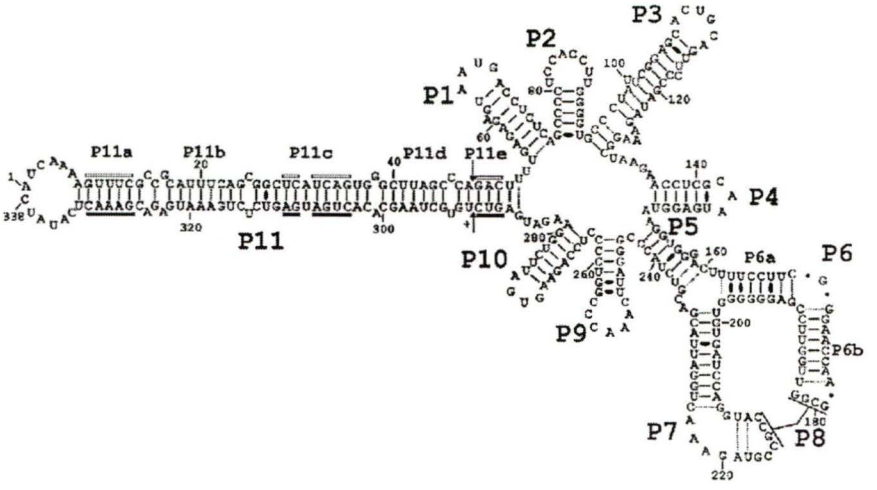


Figure 3. PLMVd sequence and the proposed secondary structure based on nuclease mapping under low salt conditions. The helix numbering has been arbritary fixed as if the circular RNA was synthesized from position 1, and follows their order of appearance. GU wobble basepairs are represented with black ovals, while Watson-Crick basepairs which appeared to be either unstable, or for which single- and double-strand coexistence is proposed under the conditions used, are illustrated with dotted lines. The basepairs of the P8 pseudoknot were not included in order to simplify the illustration.

The novel P8 pseudoknot. The proposed secondary structure for PLMVd includes a novel pseudoknot, the P8 stem. In previous secondary structures (6,7,12,13) nucleotides $_{179}$GCGG$_{182}$ (loop P6b) and $_{209}$GUACCGCCGUAGAAA$_{223}$ (loop P7) were proposed to form single stranded regions. In this scenario it is expected that the RNases with specificities for single-stranded nucleotides shoud give hydrolysis products, while RNase V_1 shoud not. In the structure proposed here, nucleotides $_{179}$GCGG$_{182}$ (loop P6b) and $_{212}$CCGC$_{215}$ (loop P7) basepair to form a pseudoknot (Fig. 4A). We observe RNase V_1 (specific for double-stranded regions) to produce several cleavage products in these regions, while both RNase T_1 (specific for single-stranded G) and T_2 (specific for single-stranded regions) were almost unable to hydrolyse the phosphodiester backbone of these bases. In addition, nucleotides $_{210}$UA$_{211}$ and $_{218}$UA$_{219}$ of the P7 loop are proposed to basepair together. These results clearly demonstrate that the P7 loop (positions 209 to 224) appears to be highly structured, with the exception of the $_{220}$GAAA$_{223}$ loop (see Table 1 and Fig. 4A).

In order to support the presence of the P8 pseudoknot, a mutant PLMVd transcript was synthesized and nuclease mapped (Fig. 4B). In contrast to the wildtype sequence, which includes the sequence $_{212}$CCGC$_{215}$ in the P7 loop, the mutant posesses the sequence $_{212}$AAAA$_{215}$. With this transcript, the RNase V_1 did not hydrolyze either nucleotides $_{212}$AAAA$_{215}$ or $_{179}$GCGG$_{182}$; whereas the four adenines at positions 212 to 215 were readily hydrolyzed by RNase T_2, and nucleotides $_{179}$GCGG$_{182}$ were readily hydrolyzed by both RNase T_2 and T_1. In addition, we observed differences in the hydrolysis patterns of nucleotides $_{210}$UA$_{211}$ and $_{218}$UA$_{219}$, suggesting a stronger base-pairing in the mutant transcript (Fig. 4B). The latter difference is probably due to the fact that the absence of the pseudoknot reduced the constraint (i.e. stress) on the P7 loop thereby favoring formation of the two basepairs. With the exception of the P8 pseudoknot region, the mutant gave a digestion pattern similar to that of the wildtype transcript.

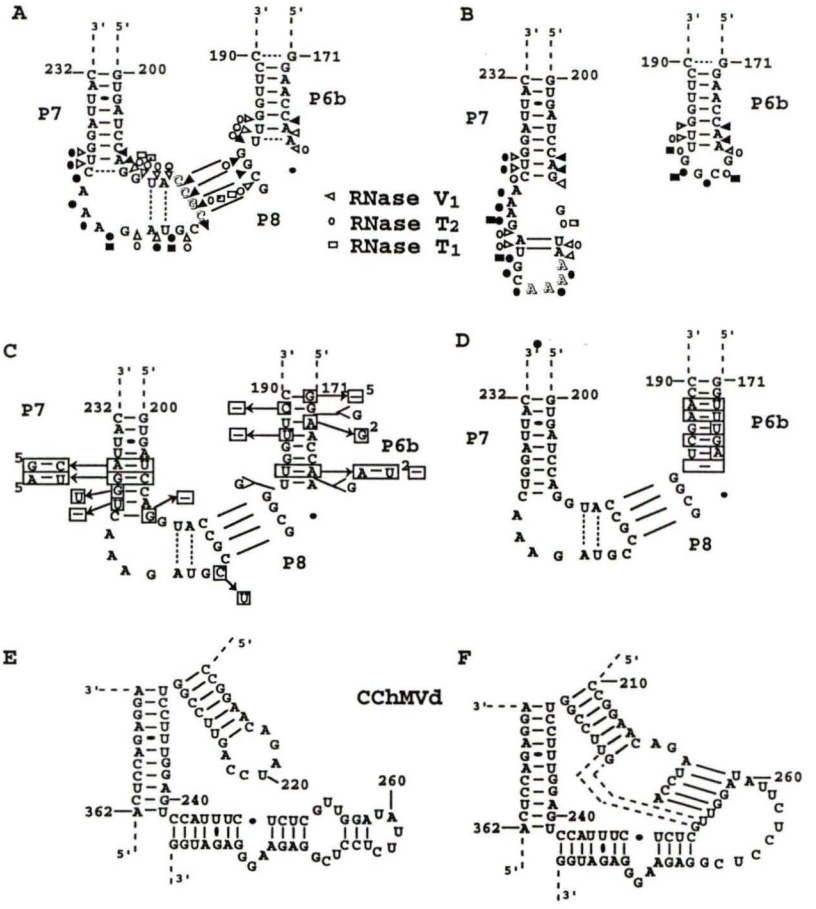


Figure 4. Characterization of the P8 pseudoknot and surrounding region. A and **B** are nuclease mapping of the wildtype ($_{212}$CCGC$_{215}$) and mutant ($_{212}$AAAA$_{215}$) PLMVd transcripts. The mutated bases are shadowed. Only results with RNases V_1, T_1 and T_2 are shown. White, dotted and black forms indicate weak, intermediate and strong cleavage, respectively. **C** and **D** are nucleotide variations observed in all known natural PLMVd isolates (14). **C** includes the sequences of 62 variants, while **D** is the sequence of a PLMVd isolated from Hardired cultivar (i.e. Hd6, see ref. 14) for which most of the P6b stem is different. Differences with the sequence characterized in this study are boxed. Uppercase numbers indicate the number of PLMVd variants which includes this mutation when it occurs more than once. **E** and **F** are the sequences and proposed secondary structures of the corresponding region of CChMVd. **E** Secondary structure as proposed previously (3). **F** Secondary structure proposed here that includes the P8 pseudoknot. Dotted lines are potential additional basepairs extending the pseudoknot.

The base composition of this four basepair pseudoknot is perfectly conserved among all PLMVd variants (Fig. 4C and D; ref. 13,14). Nucleotide variations were observed on both sides of the pseudoknot, but did not affect its length. Covariations are observed at the bottom of the P6b and P7 stems. The perfect conservation of the sequence forming the pseudoknot (i.e. four ordered GC basepairs) is probably required for formation of a helix in this stretched region of the molecule. It is also plausible that the base conservation has a biological relevance such as binding to a host macromolecule. Surprisingly, a similar pseudoknot may be formed by CChMVd, another member of the PLMVd sub-group (Figure 4E and F). According to the structural similarities between PLMVd and CChMVd (14), a pseudoknot may be formed between the P6 and P7 loops. The presence of this CChMVd pseudoknot prevents formation of 3 basepairs in the P7 stem, but it adds 5 basepairs resulting in a net of two additional basepairs which stabilize the structure (Fig. 4E and F).

Alternative structures. Based on sequence comparisons it was proposed that stem P11b and the adjacent nucleotides on both sides may adopt a slightly less stable, in terms of energy, alternative structure corresponding to the hammerhead hairpin II on both strands (Fig 5A; ref. 13). Although nuclease probing shows stem P11b to be as proposed in figure 3 for most RNA molecules, we did note that RNase Phy M hydrolyzed after the A and U in the $_{18}$AUUUCA$_{23}$ and $_{316}$UAGAAU$_{321}$ stretches, albeit at reduced levels (Table 1). One way to reconcile these results is to propose that the RNA strands adopt both structures, but strongly prefer the most stable one which allows for the stacking of all P11 sub-helices.

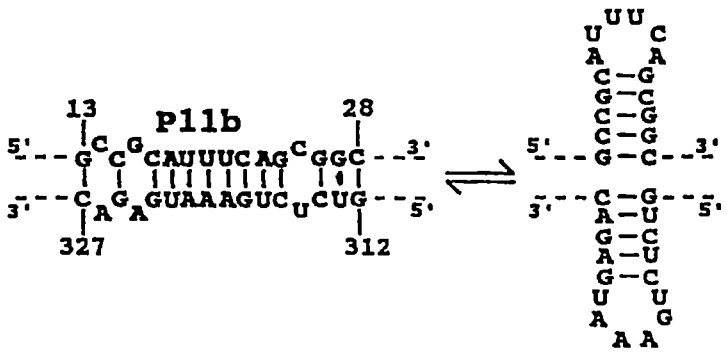


Figure 5. Alternative structure including the hairpin II of the hammerhead motifs.

The presence of a pseudoknot formed by nucleotides of the left-handed loop and the P1 loop, specifically between the sequence $_{337}$UCAU$_{2}$ and $_{65}$AUGA$_{68}$, has been suggested previously (13). While the RNA construct with the 5'-end at position 337 does not allow observation of the formation of this pseudoknot, the second construct with the 5'-end at position 245 should. The nuclease mapping results with this RNA construct suggest that both loops are located in single-stranded regions, and, therefore, the presence of this potential pseudoknot is not supported under the conditions tested (Table 1). However, sequence analysis of different PLMVd isolates shows that this pseudoknot may be extended 5 or 6 basepairs in other variants, and that the stability of this motif is related to the pathogenicity of the viroid (13). Thus it seems that this pseudoknot is not found in all PLMVd sequence variants, and consequently it maybe related to an unessential function of the viroid such as its pathogenicity (13).

Additional support from binding shift assays. Concurrent with the nuclease mapping experiments, oligonucleotide binding shift assays were performed with wildtype PLMVd transcripts starting at position 337. Trace amounts of 5'-end labelled DNA or RNA oligonucleotides were incubated in the presence of a large excess of PLMVd transcripts under low salt conditions, and the mixtures fractionated on native agarose gels. The oligonucleotides were added either before or after the transcripts were denatured at 65°C and renatured by slow cooling to 4°C. Addition of the oligonucleotide prior the denaturation-renaturation should favor binding to the transcript. Among the DNA oligonucleotides tested, trace amounts (> 0.1%) of PLMVd-oligo complexes were detected with D33-48 and D94-109 only when they were added prior to renaturation (Table 2). These two oligonucleotides correspond to one strand of the P11c-e stem and the P3 stem, respectively, and most probably hybridize to the complementary PLMVd strands. In contrast, the D39-58 oligonucleotide complementary to P11-d and -e and a portion of P1 stem produced no oligonucleotide binding, indicating that this region is tightly folded (Table

2). Oligonucleotides D153-171 and D205-224 complementary to the P5-P6 stems and the P7 stem-loop structures, respectively, hybridized with the PLMVd transcripts. Trace amounts of D153-171 was shown to hybridize with transcripts regardless of when it was added to the mixture, indicating that this region was slightly accessible to the oligonucleotide (Table 2). In contrast, D205-224 was predominantly found to hybridize when it was present during the denaturation-renaturation, whereas only trace amounts were found to hybridize when it was added after the denaturation-renaturation (Table 2). This suggests that the oligonucleotide hybridizes poorly when added to stably folded PLMVd, thereby supporting the presence of the P8 pseudoknot. The specificity of hybridization for oligonucleotides D153-171 and D205-224 was verified by RNase H hydrolysis (data not shown).

TABLE 2. Binding shift assays

oligonucleotides			hybridization[c]	
name[a]	lenght (nt)	polarity[b]	before[d]	after[d]
D33-48	16	S	+	-
D39-58	20	AS	-	-
D94-109	15	S	+	-
D153-171	19	AS	+	+
D205-224	20	AS	+++	+
R77-83	7	AS	-	-
R130-135	6	AS	++	++
R212-216	5	AS	-	-
R212-223	12	AS	++	+
R.GGG	3	AS	+++	+++

[a] R and D indicate either the RNA or DNA nature of the oligonucleotide. The numbering correspond to the position on the PLMVd sequence (see figure 3).

[b] The polarity indicates if the sequence of an oligonucleotide is sense (S) or antisense (AS; complementary) to the PLMVd transcripts.

[c] Proportion of shifted oligonucleotide: - (0), + (> 1%), ++ 1-10%, +++ >10%.

[d] Oligonucleotide were added to the mixture before or after the denaturation-renaturation of the transcripts.

Since RNA-RNA homoduplexes are more stable than DNA-RNA heteroduplexes, shorter RNA oligonucleotides were also tested in order to investigate the accessibility of short PLMVd sequences (Table 2). Oligonucleotide R77-83, which is complementary to a portion of the P2 stem-loop, constitutes a negative control, and produced no complex. These results were expected since nuclease mapping clearly indicated the presence of this stem. Oligonucleotide R130-135 hybridized efficiently to the PLMVd transcripts regardless whether it was added before or after the denaturation-renaturation, supporting the hypothesis that the small region between stems P3 and P4 is single-stranded. A short RNA oligonucleotide (R212-216, 5-mer), corresponding to the P7 loop sequence involved in formation of the P8 stem did not bind to the PLMVd transcripts, supporting the presence of the stable pseudoknot. In contrast, oligonucleotide R212-223 (12-mer) encompassing most of the P7 loop hybridized, indicating that it has the ability to unfold the P8 pseudoknot. Finally, the guanosine triplet oligonucleotide R.GGG hybridized the most efficiently of all to PLMVd transcripts under both conditions. In contrast to all other oligonucleotides tested, most likely several molecules of R.GGG hybridize to a single transcript as it has the potential to perfectly basepair with the sequences $_{81}$UCC$_{83}$, $_{85}$CCU$_{87}$, $_{86}$CUU$_{88}$ and $_{254}$CCC$_{256}$ which are all located in single-stranded regions according to nuclease mapping data. DNA and RNA oligonucleotide binding shift assays were also performed under different salt conditions (i.e. high salt) and different magnesium concentrations. With the exception of small differences in the binding affinity of some oligonucleotides, the results were similar regardless of the conditions used. Thus, the binding shift assays support the PLMVd secondary structure proposed with the data from the nuclease mapping experiments.

DISCUSSION

Structural homogeneity of the transcripts. Nuclease mapping and oligonucleotide binding shift assays confirmed that PLMVd folds into a branched secondary structure in solution. The nuclease mapping data from the two transcripts (i.e. wildtype transcripts starting either at position 377 or at position 244) used here agreed perfectly except for near the termini of the RNA molecules. Clearly both transcripts fold into a similar structure. This conclusion is supported by data from lead-induced cleavage experiments at pH 7.0 with both these transcripts and the $_{212}$AAAA$_{215}$ mutant (15). The patterns produced on PAGE gels by lead-ion hydrolysis were similar for all three RNA transcripts, clearly suggesting that they are folded in a similar manner (data not shown). The only important difference in the cleavage patterns of these three transcripts was that the $_{212}$AAAA$_{215}$ mutant's P6b and P7 loops were efficiently hydrolyzed, while those of wildtype sequence possessing the P8 pseudoknot and the basepairs $_{210}$UA$_{211}$-$_{218}$UA$_{220}$ were not. The different results presented here support the hypothesis that the PLMVd transcripts characterized in this report fold into similar structures.

PLMVd structure in solution. Complete nuclease mapping of RNA molecules longer than 100 nucleotides in length is limited to a few examples. One of these is the group B viroid PSTVd, which was unambiguously shown to fold into a rod-like structure in solution (5). In this report we use nuclease mapping, combined with oligonucleotide binding shift assays, to show that PLMVd adopts a complex, branched secondary structure in solution. In general this structure is similar to that based on sequence comparison and computer modelling (13,14), having 74.5% of its nucleotides in basepaired. Structural microheterogeneity seems to be limited to small domains including the hammerhead hairpin II and the P8 pseudoknot, the latter of which adopts primarily one structure with only a small proportion (i.e. or trace amounts) being observed to fold into alternative structures (see Results section). The use of low and high salt buffers allowed us to demonstrate that

some basepairs are stabilized in the latter conditions. For example, inside the P3 stem three basepairs were evident under high salt conditions, while under low salt conditions these basepairs seem to be formed in only a portion of the RNA molecules.

Based on the results presented here, we believe that the PLMVd structure includes the novel P8 pseudoknot. This pseudoknot also appears to be present in CChMVd, suggesting that it is a unique feature of the viroid members of the PLMVd sub-group. The structural similarities between PLMVd and CChMVd, including the presence of this pseudoknot, may help to explain why these viroids are insoluble in 2 M lithium chloride while all other viroids proposed to adopt a rod-like structure are soluble. The branched secondary structures of PLMVd and CChMVd, in conjunction with the presence of the P8 pseudoknot, produce a compact structure that is probably the biochemical basis of this observation.

The proposed structure for PLMVd is the most stable one adopted under the conditions used. However, it must be remembered that the most stable structure is not necessarily the biologically active structure. For example, hammerhead structures are crucial for self-cleavage during the rolling circle replication of PLMVd, but are not formed in the most stable PLMVd structure. Furthermore, host protein(s) may interact with PLMVd and alter its folding. Thus, a host protein may be involved in formation of the proposed pseudoknot between the P1 and P11 loops (13), and in the formation of another putative pseudoknot involving the P1 stem and the single stranded domain between the P10 and P11 stems (14), thereby explaining why these structural motifs were not observed in this structure. Conversely, the P8 pseudoknot, whose sequence is highly conserved among all PLMVd variants, is most likely to be essential in the PLMVd life cycle.

ACKNOWLEDGEMENTS

This work was sponsored by a grant from Natural Sciences and Engineering Research Council (NSERC) of Canada to J.P.P. F.B. and F.C. were the recipients of NSERC studentships. J.P.P. is a Medical Research Council (MRC) of Canada scholar.

REFERENCES

1. Flores, R., Di Serio, F. and Hernandez, C. (1997) Viroids: The noncoding genomes. *Semin.Virol.*, **8**, 65-73.
2. Symons, R.H. (1997) Plant pathogenic RNAs and RNA catalysis. *Nucleic Acids Res.*, **25**, 2683-2689.
3. Navarro, B. and Flores, R. (1997) Chrysanthemum chlorotic mottle viroid: unusual structural properties of a subgroup of self-cleaving viroids with hammerhead ribozymes. *Proc. Natl. Acad. Sci. USA*, **94**, 11262-11267.
4. Bussière, F., Lafontaine, D. and Perreault, J.P. (1995) Compilation and analysis of viroid and viroid-like RNA sequences. *Nucleic Acids Res.*, **24**, 1793- 1798.
5. Gast, F.U., Kempe, D., Spieker, R.L. and Sänger, H.L. (1996) Secondary structure probing of potato spindle tuber viroid (PSTVd) and sequence comparison with other small pathogenic RNA replicaons provides evidence for central non-canonical base-pairs, large A-rich loops, and a terminal branch. *J. Mol. Biol.*, **262**, 652-670.
6. Hernandez, C. and FLores, R. (1992) Plus and minus RNAs of peach latent mosaic viroid self-cleave *in vitro* via hammerhead structures. *Proc. Natl. Acad. Sci. USA*, **89**, 3711-3715.
7. Beaudry, D., Bussière, F., Lareau, F., Lessard, C. and Perreault, J.P. (1995) The RNA of both polarities of the peach latent mosaic viroid self-cleaves in *vitro* solely by single hammerhead structures. *Nucleic Acids Res.*, **23**, 745-752.

8. Beaudry, D. and Perreault, J.P. (1995) An efficient strategy for the synthesis of circular RNA molecules. *Nucleic Acids Res.*, **23,** 3064-3066.

9. Zawadski, V., and Gross, H.J. (1991) Rapid and simple purification of T7 RNA polymerase. *Nucleic Acids Res.*, **19**, 1948.

10. Davanoo, P., Rosenberg, A.H., Dunn, J.J. and Studier, F.W. (1984) Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl. Acad. Sci. USA*, **81**, 2035-2039.

11. Perreault, J.P. and Altman, S. (1992) Important 2'-hydroxyl groups in model substrates for M1 RNA, the catalytic subunits of RNase P from *E. coli. J. Mol. Biol.*, **226**, 399-409.

12. Bussière, F., Lehoux, J., Thompson, D.A., Skrzeczkowski, L.J. and Perreault, J.P. (1999) Subcellular localization and rolling circle replication of peach latent mosaic viroid: Hallmarks of group A viroids. *J. Virol.*, **73**, in press.

13. Ambros, S., Hernandez, C., Desvignes, J.C. and Flores, R. (1998) Genomic structure of three phenotypically different isolates of peach latent mosaic viroid: Implication of the existence of constraints limting the heterogeneity of viroid quasispecies. *J. Virol.*, **72**, 7397-7406.

14. Pelchat, M., Lévesque, J. Ouellet, D. Laurandeau, S., Lévesque, S., Lehoux, J., Thompson, D.A., Skrzeczkowski, L.J. and Perreault, J.P. (1999) Sequencing of peach latent mosaic viroid variants from nine peach cultivars supports a complex secondary structure that include three putative pseudoknots. submitted to *J. Virol.*

15. Ouellet, J. and J.P. Perreault. (1999) unpublished data.

Description du travail

L'objectif principal du travail était d'étudier les éléments structuraux de PLMVd en solution. En utilisant l'approche de susceptibilité aux ribonucléases nous avons caractérisé la structure secondaire du viroïde. L'utilisation de deux constructions linéarisées à des endroits différents a permis d'étudier la molécule dans sa totalité. Les résultats montrent l'existence d'une structure tertiaire (pseudo-noeud) qui est confirmée par la caractérisation d'un mutant incapable de former cette intercation. De plus, plusieurs caractéristiques structurales de PLMVd semble être retrouvées chez CChMVd, un viroïde du groupe A récemment identifié. On observe donc pour la première fois une organisation structurale potentiellement conservée chez des viroïdes du groupe A. PLMVd et PSTVd (groupe B) sont actuellement les deux seuls viroïdes pour lesquels on bénéficie d'une telle connaissance structurale. Le manuscrit est dans sa forme finale et a été récemment soumis, pour publication, à la revue *Journal of Virology*.

Implication dans le travail

J'ai mis au point les conditions à utiliser pour la réalisation des expériences de susceptibilité aux ribonucléases que j'ai ensuite appliqué pour caractériser une bonne partie de la structure secondaire de PLMVd. Les expériences ont été complétées par les co-auteurs de l'article. J'ai également proposé les stratégies à utiliser pour la réalisation des différentes constructions, dont celle du mutant permettant de prouver l'existence du pseudo-noeud.

CHAPITRE 4:

Compilation des séquences représentant les viroïdes et ARN apparentés

4.1 ARTICLE:

Compilation and analysis of viroid and viroid-like RNA sequences

Frédéric Bussière, Daniel Lafontaine et Jean-Pierre Perreault

Département de Biochimie

Université de Sherbrooke

Sherbrooke, Québec, J1H 5N4, Canada

Article publié dans: *Nucleic Acids Research*, Vol. 24, 1793-1798, 1996.

ABSTRACT

We have created a catalogue comprising all viroid and viroid-like RNA sequences which to our knowledge have been either published or were available from on-line sequence librairies as of October 1, 1995. In the development of this catalogue nomenclature ambiguities were removed, the likely ancestral sequence of most species was determined and the most stable secondary structures of these sequences were predicted using the Mulfold parkage. Only viroids of PSTVd-type possessed a rod-like secondary structure, while most other viroids adopted branched secondary structures. Several viroids have predicted secondary structures that include either a Y or cruciform structure reminiscent of the tRNA-like end of virus genomes at an extremity. However, it remains unknown whether or not these predicted structures are adopted in solution, and if they serve a particular function *in vivo*. Additional information such as the position of the self-catalytic domains are included in the catalogue. An analysis of the data compilated in the catalogue is included. The catalogue will be available on the world wide web (http://www.callisto.si.usherb.ca/~jpperra), computer disk and in printed form. It should provide an excellent reference point for further studies.

INTRODUCTION

Viroids are small single-stranded circular RNA molecules (246-463 nucleotides) that infect higher plants, causing diseases in crop species and resulting in important economic losses in the agricultural industry (1). It has been proposed that viroids replicate in a DNA-independent manner via a rolling circle mechanism involving the synthesis of multimeric strands which are then cleaved into monomeric fragments and circularized producing the progeny viroids (1,2). We have developped a catalogue in order to facilitate viroid research by presenting a large amount of viroid sequence and related data in a comprehensive and user-friendly format. We compilated a total of 182 sequences from 21 viroids (3-79), 8 plant satellite viroid-like RNAs (80-88) and the viroid-like domain of the human hepatitis d virus RNA (89). In addition, we have included a transcript from the mitochondrial DNA satellite II from both *newt* (90) and carnation (91) which have been proposed to be retroviroid-like elements evolving from viroids by retroposition into host DNA (92). Table 1 is a summary of these sequences. Here, we describe the catalogue and present an analysis of its content.

THE CATALOGUE

This compilation comprises all sequences that to our knowledge had been published or were available from the sequence library file servers (NCBI sequence librairies) as of October 1st, 1995. Only complete sequences were retained because it appeared difficult to establish a criteria on how large a sequence fraction should be in order for it to be listed in the compilation. With the heterogeneity that occurs in viroids, the inclusion of sequences from partial cDNA clones could result in erronious conclusions in subsequent studies, and therefore they were omitted.

RNA species	Abbreviation	number of compilated sequences	size (nt)	selected variants	size of selected variant (nt)	catalytic motifs
Viroids						
ASBVd-type (group A)						
avocado sunblotch viroid	ASBVd	19	246-251	ASBVd.2	247	hh+ / hh-
peach latent mosaic viroid	PLMVd	2	337-338	PLMVd.1/.2	337-338	hh+ / hh-
PSTVd-type (group B)						
PSTVd group (subgroup B1)						
coconut cadang-cadang viroid	CCCVd	7	246-301	CCCVd.2	246	
citrus exocortis viroid	CEVd	31	368-463	CEVd.16	371	
columnea latent viroid	CLVd	2	370-372	CLVd.1	370	
chrysanthemum stunt viroid	CSVd	2	354-356	CSVd.2	354	
coconut tinangaja viroid	CTiVd	2	254	CTiVd.1-.2	254	
citrus viroid species IV	CVd IV	1	284	CVd IV.1	284	
hop latent viroid	HLVd	1	256	HLVd.1	256	
hop stunt viroid	HSVd	23	294-303	HSVh.1	297	
potato spindle tuber viroid	PSTVd	27	341-361	PSTVd.1	359	
tomato apical stunt viroid	TASVd	2	360-363	TASVd.1/.2	360	
tomato planta macho viroid	TPMVd	1	360	TPMVd.1	360	
ASSVd group (subgroup B2)						
australian grapevine viroid	AGVd	1	369	AGVd.1	369	
apple scar skin viroid	ASSVd	4	329-331	ASSVd.3	329	
citrus bent leaf viroid	CBLVd	2	315-318	CBLVd.2	315	
citrus viroid species III	CVd-III	2	294-297	CVd-III.2	294	
coleus viroid	CoVd	2	248	CoVd.1/.2	248	
grapevine yellow speckle viroid	GYSVd	33	363-369	GYSVd.27	368	
grapevine 1B viroid	G1BVd	1	363	GV1Bd.1	363	
pear blister canker viroid	PBCVd	1	315	PBCVd.1	315	
Satellite RNA						
Luteovirus						
barley yellow dwarf virus satellite RNA	vBYDV	1	322	vBYDVd.1	322	hh+ modified / hh-
Nepovirus						
arabis mosaic virus satellite RNA	sARMV	1	300	sARMV.1	300	hh+ / hp-
chicory yellow mottle virus satellite RNA S1	sCYMV-S1	1	457	sCYMV-S1.1	457	hh+ / hp-
lucerne transient streak virus satellite RNA	vLTSV	3	322-324	vLTSV.3	322	hh+ / hh-
tobacco ringspot virus satellite RNA	sTobRV	2	359-360	sTobRV.1	359	hh+ / hp-
Sobemovirus						
subterranean clover mottle virus satellite RNA	vSCMoV	2	332-338	vSCMoV.2	332	hh+
solanum nodiflorum mottle virus satellite RNA	vSNMV	1	377	vSNMV.1	377	hh+
velvet tobacco mottle virus satellite RNA	vVTMoV	2	365-366	vVTMoV.1	365	hh+
Other related RNA						
carnation stunt associated viroid	rCarSV	2	275	CarSV.1/.2	275	hh+ / hh-
Hepatitis delta virus*	vHDV	1	367	vHDV	367	delta+ / delta-
Newt satellite 2 transcript	rNS2T	1	281	Newt.1	281	hh+

TABLE 1. Summary of the viroid and viroid-like RNA sequences included in the catalogue. The selected variants are the likely ancestral ones determined as described in the text. The presence of known catalytic motifs are identified by hh for hammerhead, hp for hairpin and delta for self-catalytic domains found in the hepatitus delta virus. "+" and "-" indicate the polarity of the RNA which possesses the self-catalytic domain. "d" indicates a viroid, "v" a virusoid (i.e. a plant viroid-like satellite circular RNA), "s" a plant viroid-like linear satellite RNA and "r" a viroid-like retroelement. "nt" is for nucleotide. *Because it is only a domain of HDV that is related to the viroid, only one variant sequence has been included even thougth several other sequences have been determined.

Among viroids, the classification system proposed by Koltunow and Rezaian has been adopted (93). Viroids are divided in two types, the ASBVd-type (also named group A) whose members possess the capacity to self-cleave, but do not possess a conserved core region (CCR); and, the PSTVd-type (or group B) whose members possess a CCR, but have no known self-cleaving properties. Viroids of PSTVd-type are subdivided in two groups: the PSTVd group (or subgroup B1) and the ASSVd group (or subgroup B2). The viroids of these two groups are easily distinguished by the sequence forming their respective CCRs (93). Only the sequences of HSVd, which is a member of PSTVd group, were subdivided according to their initial isolation host.

In the catalogue each species is listed by its complete name and number of sequence variants (see Fig.1). This is followed by, for each species, a listing of the sequence variants with their new identifications (see below), accession numbers for sequence library file server, bank loci (when available), number of nucleotides, numbers of each type of nucleotide, complete publication information, and the complete nucleotide sequence in 10 nt blocks in order to facilitate further analyses. For species which possess known self-catalytic domains (hammerhead, hairpin and delta), the localization of the conserved sequences required or cleavage to occur are reported. In addition, a secondary structure prediction of the most likely ancestral variant was derived using MulFold structure prediction package (see below). The chosen most likely ancestral variants are reported in Table 1, while their predicted secondary structures are appended to the catalogue.

ASSVd-type (subgroup B2)
- Australian grapevine viroid (AGVd), 1 variant

>AGVd.1
accession numbers:X17101(embl), 58574(gi)
Locus Agvccs
369 nucleotides (76A, 103C, 111G, 79T)
Rezaian,M.A.Australian grapevine viroid-evidence for extensive recombination.Nucleic Acids Res. 18, 1813-1818 (1990)

```
TGGGCACCAA CTAGAGGTTC CTGTGGTACT CACCGAAGGC CGCGAACGTA GGAAAGAAAA
AGATAGAAAA GCTGGGTAAG ACTCACCTGG CGACTCGTCG TCGACGAAGG GTCCTCAGCA
GAGCACCGGC AGGAGGCGCT ATGCAGGAAC GCTAGGGGTC CTCCAGCGGA GGACTGAAGA
AACTCCGGTT TCTTCTTTCA CTCTGTAGCT GGAATCCCTG TTGCGCTTGC TGGCGAAACC
TGCAGGGAAG CTAGCTGGGT CCCGCTAGTC GAGCGGACTC GTCCCAGCGG TCCCAACCAG
TTTTCTTTAT CCTATTTTTC CTGCGGGCGC CCGGTCGTGG TTACCCTGGA GCTCCCTGTT
TGGAGGCCC
```

Figure 1: Example of the AGVd sequence catalogue entry.

PROPOSITION OF AN ACCURATE IDENTIFICATION SCHEME

In order to simplify the nomenclature of the included sequences, we used an identification scheme based on the usual abbreviation of an RNA species followed by a number. For example, the original PSTVd sequence (30) is identified as PSTVd.1. The number is function of the date of report. For sequences already published or reported in on-line libraries, priority was given to publication date over library submission date. When more than one sequence was reported simultaneously, we attributed arbitrary numbers to the entries. Exclusively for HSVd variants, a letter according to the initial isolate host precedes the number. For example, HSVd.h1 refers to a species isolate from the hop. We believe that the proposed nomenclature will facilitate further sequence identification.

In creating this catalogue we clarified some identification ambiguities. Our proposed clarifications are listed below:

(i-) Previously, two sequences have been identified as grapevine viroid, GVVd (7,25). Phylogenetic analysis of these sequences (data not shown) clearly demonstrated that one is a variant of CEVd (7), which we identified as CEVd.22; while the other is a HSVd variant infecting the grapevine (25), which we identified as HSVd.g5;

(ii-) The second sequence reported as a GYSVd variant has been reported elsewhere as a distinct species, GV1B (68). We used the latter identification as it is supported by phylogenetic analysis which assigned 52 substitutions between the GVIB and the likely ancestral GYSVd.27 variant (data not shown);

(iii-) Two sequence variants of CPFVd have been reported (28,29). They belong to HSVd species isolated from cucumber. We have identified these two sequences as HSVd.c1 and .c2.

(iv-) Dapple apple viroid (DAVd; EMBL acession number X71599) appeared as the likely ancestral variant of apple scar skin viroid (ASSVd.4), and not as a different species.

(v-) We used the nomenclature CoVd (*Coleus viroids*) which included the *Coleus blumei viroid species I* (CoVd.1, ref. 64) and the *Coleus yellow viroid* (CoVd.2, ref. 65). This is in agreement with the suggestion of Fonseca and al. (65).

ANALYSIS OF THE CATALOGUE

We analyzed the compilated sequences for the general characteristics of viroids and viroid-like RNAs taking for account the phylogenetic reconstruction reported by Elena et al. (94), as well as an updated reconstruction confirming the previous one (unpublished data, F. Bussière, D. Lafontaine and J.-P. Perreault). No relationship can be discerned between the nucleotide percentages, sequence length and phylogenetic clustering. The nucleotide differences between variants of the same species are located primarily in the pathogenesis

(P) and the variable (V) domains of the viroids, however they are not restricted to these two domains as concluded previously (95). Furthermore, no relationship has been established between viroids and either their host or their worldwide geographic distribution, nor have any specific signature sequence been identified for viroids which infect the same host. In contrast, some interesting features were pointed out by phylogenetic reconstructions and secondary structure predictions.

Prediction of the most likely ancestral variants

In order to study the relationship between variants belonging to the same species and to identify the most likely ancestral sequence, we performed either direct inferences or phylogenetic reconstructions. The identification of the probable ancestral variants will be useful in further phylogenetic reconstructions between species. When two sequences were known, we initially looked to see if one could be derived from the other, and then used the ancestral one if this was indeed the case. If they did not fulfill this requirement, the sequences were considered as belonging to different taxa. If more than two variants existed, the ancestral state was either directly inferred, or obtained by phylogenetic analysis. In the latter cases the sequence alignments were carried out by means of a multiple sequence algorithm with hierchical clustering (Multalin package, ref. 96) followed by minor sequence rearrangements, and then the phylogenetic analysis was performed using maximum parsimony method (PAUP package, ref. 97). We deduced the current probable ancestral variant among the sequences available in table 1, and appended this list to the compilation. The addition of more sequences may facilitate the identification of ancestral variants.

Exhaustive phylogenetic reconstitutions of the species that have a large number of variants (ASBVd, CCCVd, CEVd, GYSVd, HSVd and PSTVd) has been performed. The case of CCCVd was the simplest among these viroids. Five CCCVd variants resulted from a duplication of a region of CCCVd.2 and then subsequent substitutions occurred (CCCVd.3-CCCVd.7), as suggested previously (95). CCCVd.1 has a one nucleotide

change as compared to CCCVd.2, and this substitution is not found in the other five larger variants; therefore, we infered CCCVd.2 as the oldest variant based on the available data. In contrast, the other viroids with several variants required phylogenetic reconstructions to identify the likely oldest sequence. Most of the variants of these viroids differ only by a very small number of substitutions (~1-10 mutations and/or deletions and/or insertions). When all variants of one of these species were considered for a phylogenetic reconstruction, several sequences could have evolved in different manners from either the ancestral variants or other variants. Thus, several trees with the same total length were inferred for a specific species (also observed by Dr. Robert Owens for HSVd and PSTVd, personal communication). Because no biological data permits validation of these phylogenetic trees (for example: replicational data), a concensus tree was deduced and the likely oldest sequence defined as the ancestral one.

Possibly the identity of the host infected by a viroid may be useful for the validation of phylogenetic trees. For example HSVd variants infect a wide range of hosts including hop, grapevine, plum, pear, peach, citrus and cucumber. Taking for account the host specificity of HSVd variants, we analyzed the trees derived in order to verify whether variant clusters reflected their host specificity. As previously reported (26), the HSVd variants clustered mainly in three variant types: the hop (hop, grapevine, peach and pear isolates), the plum (grapevine, peach and plum isolates) and the citrus (citrus and cucumber isolates). The HSVd phylogenetic tree did not strictly reflect the host specificity; hence this did not allow validation of the phylogenetic data.

For viroids that have only two known variants, the sequences are nearly perfectly identical, and therefore did not allow identification of the ancestor. One exception was TASVd in which TASVd.1 and .2 show a sequence homology of 91.5% (50) and appear the most distant variants of the same species with two known sequences. A phylogenetic reconstruction of some PSTVd-type viroids shows that the two TASVd sequences branched onto the main lineage (Fig. 2). These two branches were separated by an average of 16

substitutions. These results may support two different possibilities, either the two TASVd sequences belong to two different species which infect the same host, or they are two variants of a same species. The latter case suggests that CEVd has evolved directly from TASVd and not from a common ancestor. Further characterization should determine which of these two scenarios is the correct one.

Secondary structure predictions

In order to allow comparison between the most stable secondary structures of the ancestral species, we performed computer analysis on the previously selected sequences. We used the Mulfold structure prediction package (98) of GCG (Genetic Computing Group) version 8.0 installed on a UNIX system (at the Institut de Recherche Clinique de Montréal). For each selected sequence, a prediction of the secondary structure was obtained, and the resulting structures transformed into "connect" files using the plotfold-H software. The connect file of each predicted structure follows the sequence in the website, thereby allowing any investigator to work within it using his own graphics package. For the printed version of the catalogue, the secondary structure connect files have been displayed using the RNADrawn package and are appended.

Most of the published viroid secondary structures were predicted with the original RNAfold package. These original predictions led to proposal that the most stable secondary structure of the classical viroids of the PSTVd-type (i.e. PSTVd and ASSVd groups) are rod-like structures composed of alternating single- and double-stranded regions (1,95). With time, the rod-like secondary structure almost became an identification criterion for viroids. Analysis of our predicted secondary structures led to unexpected results that are summarized in figure 3. To facilitate the analysis, the results are presented based on the phylogenetic tree of viroids and related RNAs.

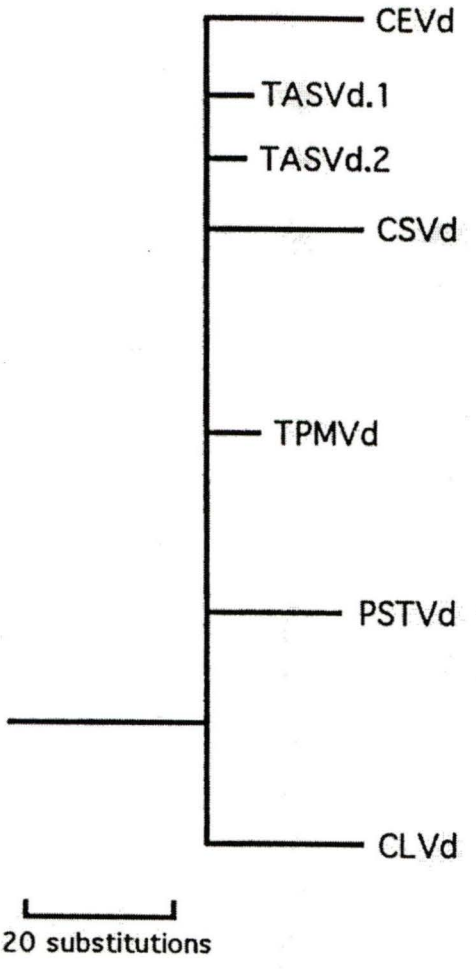


Figure 2. Phylogenetic tree of some PSTVd group viroids.

Excluding CEVd and CSVd, all members of the PSTVd group (including HSVd, CCCVd and related viroids) adopted a rod-like most stable secondary structure. In contrast, CEVd and CSV have at one end of their rod-like structure a cruciform tRNA-like structure. The CEVd cruciform structure involves 80 nucleotides at the left extremity, while the CSVd version is composed of 73 nucleotides and is located at the right end of the rod. These cruciform structures resemble the tRNA-like structure of RNA virus genomes. However, the tRNA-like structure predicted in these two species have not been demonstrated to occur in solution, nor they have a function attributed to them.

The structure predictions of the viroids from ASSVd group show a gradual evolution from rod-like secondary structure to more and more branched structures. AGVd is a perfect rod-like shape, ASSVd adopts a rod-like secondary structure with a small Y at the left end, while GYSVd has several branches leading to the absence of the initial basic rod-like structure. Other viroids of the ASSVd group follow the same scheme as a function of their phylogenetic position.

ASBVd-type viroids are ASBVd and PLMVd. ASBVd adopts a mostly rod-like secondary structure with a small Y at the left end. In contrast, both PLMVd variants adopt branched secondary structures. Similarly, most of the viroid-like satellite RNAs adopted branched secondary structures which included the basic rod-like shape and a tRNA-like structure at one extremity. Finally, the two transcripts related to viroids, rCarSV and rNS2T, and the only luteovirus (vBYDV) adopted branched secondary structures, while the viroid-like domain vHDV formed a long rod-like structure (data not shown).

The case of the two PLMVd sequence variants established by Hernandez and Flores (78), led to some interesting observations. The most stable structure of PLMVd.1 was originally predicted using the RNAfold package (78), and has recently been revised by them (personal communication) using the Mulfold package resulting in several minor differences. We have confirmed their latest prediction. The left hand domain, which is composed of the sequences involved in the formation of the hammerhead motifs of both polarities, is rod-like

in structure (78). Eleven of the 15 nucleotide differences between the two PLMVd variants are in the left hand region, which is the most stable region of PLMVd (79). These nucleotide differences affect the native secondary structure only locally. In contrast, the central and right hand regions are composed of several hairpins and 4 out of the 15 nucleotide differences are sufficient to yield relatively different structures between the variants. These results clearly show that minimal nucleotide differences may drastically affect the predicted most stable secondary structure; hence, this may be important for the biological activity associated with these regions. Analysis of the computer predicted secondary structures is instructive for the formulation of hypotheses on structure/function relationships. However, *in vitro* as well as *in vivo* characterization of biological structures is obviously more accurate for the structure/function relationship.

Analysis of both the free energy values and the proportion of GC, AU and GU base-paired regions of the predicted secondary structures as a function of either clustering or placing in a phylogenetic tree did not lead to any conclusions. Moreover, analysis of the position of the nucleotide differences does not indicate that co-variations are frequently observed within viroid sequences. However, this may due to our lack of knowledge of viroid biological secondary structures. This affirmation is supported by the analysis of the hammerhead sequence present in PLMVd. As mentionned previously, the eleven nucleotide differences between the two PLMVd variants in the left hand region affect the most stable secondary structure only locally. However, these nucleotide differences do not affect the hammerhead secondary structures (79). As a result co-variation of base paired nucleotides is observed, suggesting a selective pressure in favor of the self-cleavage activity (79). This example shows that the co-variation is associated with the secondary structures that have biological importance. Clearly, the identification of the viroid biological secondary and tertiary structures is an important research avenue.

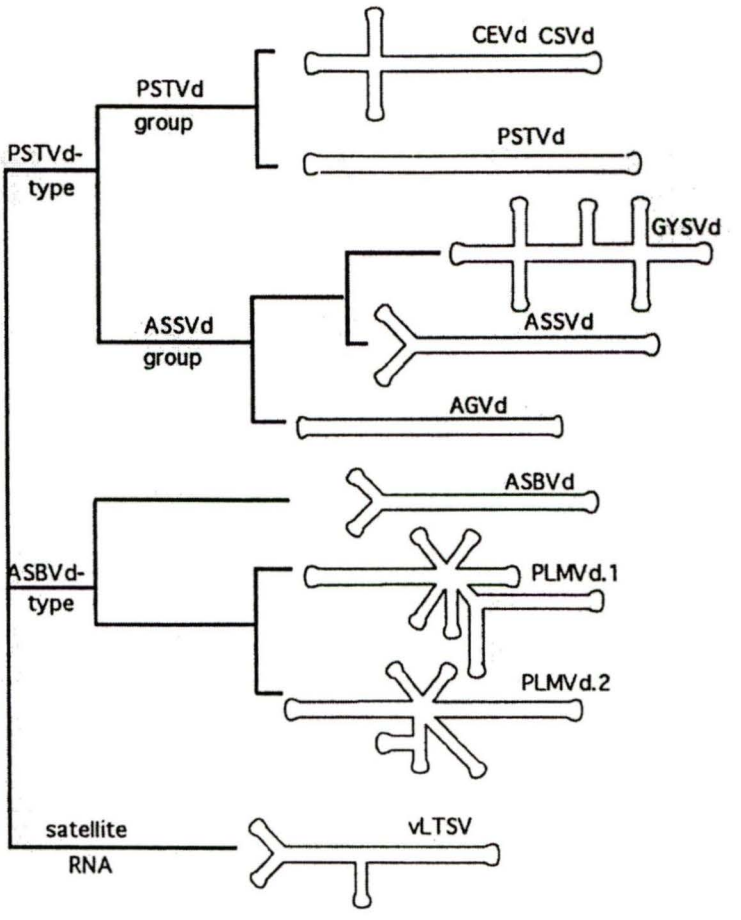


Figure 3. Schematic representation of the predicted secondary structure derived using the Mulfold structure prediction package and considering the phylogenetic relationship. The length of the lineages are not proportional to the number of substitutions between the species.

COMPLETLENESS, ACCURACY AND AVAILABILITY OF THE DATA

We have attempted to compile a catalogue of all viroid and viroid-like sequences published in journals or available from the GenBank and EMBL nucleotide sequence librairies. To this end, the EMBL and NCBI library file servers were scanned using several queries for new sequences. Sequences obtained were added to the catalogue as well as any pertinent, useful information. The authors would appreciate being informed of any omitted sequences or errors in the data set. We intend to correct any such errors in the future. Furthermore, the catalogue will be updated a few times per year. In future catalogue updates, no priority will be attributed to publication. This compilation will be available on the world wide web (http://www.callisto.si.usherb.ca/~jpperra). In this manner all viroid researchers will have the possiblity, as well as some responsibility, of updating this compilation by electronic mail submission (jp.perre@courrier.usherb.ca) following the example in figure 1. The viroid and viroid-like RNA catalogue will be also available on floppy disks, readable on microcomputers operating under MS-DOS, or in hard copy form.

ACKNOWLEDGEMENTS

The authors thank Drs. Rudra P. Singh and Robert A. Owens for critical review of sections of the catalogue. This work was supported in part by grants from the Natural Sciences and Engineering Research Council (NSERC) and by the Medical Research Council (MRC) of Canada to J.-P.P, and in part by a scientific team grant from Fonds pour la Formation des Chercheurs et l'Avancement de la Recherche du Québec (FCAR). F.B. and D.L. are recipients of predoctoral fellowships from FCAR and Fonds de la Recherche en Santé du Québec (FRSQ), respectively. J.-P.P. holds a schorlarship from the Medical Research Council of Canada.

REFERENCES

1 Diener,T.O. (1993) Trends Microbiol., 1, 289-294.

2 Branch,A.D. and Robertson,H.D. (1984) *Science*, **223**, 450-455.

3 Gross,H.J., Krupp,G., Domdey,H., Raba,M., Jank,P., Lossow,C., Alberty,H., Ramm,K. and Sänger,H.L. (1982) *Eur. J. Biochem.*, **121**, 249-257.

4 Visvader,J.E., Gould,A.R., Bruening,G.E. and Symons,R.H. (1982) *FEBS Lett.*, **137**, 288-292.

5 Visvader,J.E. and Symons,R.H. (1983) *Virology*, **130**, 232-237.

6 Visvader,J.E. and Symons,R.H. (1985) *Nucleic Acids Res.*, **13**, 2907-2920.

7 Garcia-Arenal,F., Pallas,V. and Flores,R. (1987) *Nucleic Acids Res.*, **15**, 4203-4210.

8 Mishra,M.D., Hammond,R.W., Owens,R.A., Smith,D.R. and Diener,T.O. (1991) *J. Gen. Virol.*, **72**, 1781-1785.

9 Semancik,J.S., Szychowski,J.A., Rakowski,A.G. and Symons,R.H. (1993) *J. Gen. Virol.*, **74**, 2427-2436.

10 Haseloff,J., Mohamed,N.A. and Symons,R.H. (1982) *Nature*, **299**, 316-321.

11 Rodriguez, M.J.B. and Randles, J.W. (1993) *Nucleic Acids Res.*, **21**, 2771.

12 Hammond,R., Smith,D.R. and Diener,T.O. (1989) *Nucleic Acids Res.*, **17**, 10083-10094.

13 Singh,R.P., Lakshman,D.K., Boucher,A. and Tavantzis,S.M. (1992) *J. Gen. Virol.*, **73**, 2769-2774.

14 Haseloff,J. and Symons,R.H. (1981) *Nucleic Acids Res.*, **9**, 2741-2752.

15 Gross,H.J., Krupp,G., Domdey,H., Raba,M., Jank,P., Lossow,C., Alberty,H., Ramm,K. and Sänger,H.L. (1982) *Eur. J. Biochem.*, **121**, 249-257.

16 Keese,P., Osorio-Keese,M.E. and Symons,R.H. (1988) *Virology*, **162**, 508-510.

17 Puchta,H., Ramm,K., Luckinger,R., Hadas,R., Bar-Joseph,M. and Sänger,H.L. (1991) *Nucleic Acids Res.*, **19**, 6640-6640.

18 Puchta H., Ramm K. and Sänger H.L. (1988) *Nucleic Acid Res.*, **16**, 4197-4216.

19 Ohno,T., Takamatsu,N., Meshi,T. and Okada,Y. (1983) *Nucleic Acids Res.*, **11**, 6185-6197.

20 Lee,J.Y., Puchta,H., Ramm,K. and Sänger,H.L. (1988) *Nucleic Acids Res.*, **16**, 8708-8708.

21 Sano,T., Hataya,T., Terai,Y. and Shikata,E. (1989) *J. Gen. Virol.*, **70**, 1311-1319.

22 Puchta,H., Ramm,K., Luckinger,R., Freimueller,K. and Sänger,H.L. (1989) *Nucleic Acids Res.*, **17**, 5841-5841.

23 Sänger,H.L. (1989) *Nucleic Acids Res.*, **17**, 1247-1247.

24 Puchta,H., Ramm,K. and Sänger,H.L. (1988) *Nucleic Acids Res.*, **16**, 2730-2730.

25 Sano,T., Ohshima,K., Hataya,T., Uyeda,I., Shikata,E., Chou,T.-G., Meshi,T. and Okada,Y. (1986) *J. Gen. Virol.*, 67, 1673-1678.

26 Hsu,Y., Chen,W. and Owens,R.A. (1995) *Virus Genes*, **9**, 193-195.

27 Sano,T., Hataya,T., Terai,Y. and Shikata,E. (1989) *J. Gen. Virol.*, **70**, 1311-1319.

28 Puchta,H., Ramm,K. and Sänger,H.L. (1988) *Nucleic Acids Res.*, 16, 8171-8171.

29 Sano,T., Uyeda,I., Shikata,E., Ohno,T. and Okada,Y. (1984) *Nucleic Acids Res.*, **12**, 3427-3434.

30 Gross,H.J., Domdey,H., Lossow,C., Jank,P., Raba,M., Alberty,H. and Sänger,H.L. (1978) *Nature*, **273**, 203-208.

31 Hammond,R.W. and Owens,R.A. (1987) *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 3967-3971.

32 Diener,T.O. (1979) *Science,* **205**, 859-866.

33 Kuo,T.-T. (1979) *NTU Phytopathol. Enzymol.*, **6**, 1-9.

34 Gross,H.J. and Riesner,D. (1980) *Angew. Chem. Int. Ed. Engl.*, **19**, 231-243.

35 Gross,H.J., Liebl,U., Alberty,H., Krupp,G., Domdey,H., Ramm,K. and Saenger,H.L. (1981) *Biosci. Rep.*, **1**, 235-241.

36 Singh,R.P., Singh,M., Boucher,A. and Owens,R.A. (1993)*Canadian journal of plant pathology*, **15**, 134-138.

37 van Wezenbeek,P., Vos,P., van Boom,J.H. and van Kammen,A. (1982) *Nucleic Acids Res.*, **10**, 7947-7957.

38 Schnoelzer,M., Haas,B., Ramm,K., Wang,Z.F. and Sänger,H.L. (1985) *EMBO J.*, **4**, 2181-2190.

39 Tsagris,M., Tabler,M. and Saenger,H.L. (1991) *Nucleic Acids Res.*, **19**, 1605-1612.

40 Steger,G., Baumstark,T., Moerchen,M., Tabler,M., Tsagris,M., Sänger,H.L. and Riesner,D. (1992) *J. Mol. Biol.*, **227**, 719-737.

41 Qu,F., Heinrich,C., Loss,P., Steger,G., Tien,P. and Riesner,D. (1993) *EMBO J.*, **12**, 2129-2139.

42 Owens,R.A.,Khurana,S.M.,Smith,D.R.,Singh,M.R. and Garg,I.D. (1992) *Plant Disease*, **76**, 527-529.

43 Puchta,H., Herold,T., Verhoeven,K., Roenhorst,A., Ramm,K., Schmidt-Puchta,W. and Sänger,H.L. (1990) *Plant Mol. Biol.*, **15**, 509-511.

44 Herold,T., Haas,B., Singh,R.P., Boucher,A. and Sänger,H.L. (1992) *Plant Mol. Biol.*, **19**, 329-333.

45 Lakshman,D.K. and Tavantzis,S.M. (1993) *Arch. Virol.*, **128**, 319-331.

46 Gora,A., Candresse,T. and Zagorski,W. (1994) *Arch. Virol.*, **138**, 233-245.

47 Wassenegger,M., Heimes,S. and Sänger H.L. (1994) *EMBO J.*, **13**, 6172-6177.

48 Gruner,R., Fels,A., Qu,F., Zimmat,R., Steger,G. and Riesner D. (1995) *Virology*, in press.

49 Kiefer,M.C., Owens,R.A. and Diener,T.O. (1983) *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 6234-6238.

50 Candresse,T., Smith,D. and Diener,T.O. (1987) *Nucleic Acids Res.*, **15**, 10597-10597.

60 Rezaian,M.A. (1990) *Nucleic Acids Res.*, **18**, 1813-1818.

61 Hashimoto,J. and Koganezawa,H. (1987) *Nucleic Acids Res.*, **15**, 7045-7051.

62 Puchta,H., Luckinger,R., Yang,X., Hadidi,A. and Sänger,H.L. (1990) *Plant Mol. Biol.*, **14**, 1065-1067.

63 Ashulin,L., Lachman,O., Hadas,R. and Bar-Joseph,M. (1991) *Nucleic Acids Res.*, **19**, 4767-4767 (1991).

64 Spieker,R.L., Haas Charng,Y.C., Freimuller,K. and Sänger,H.L. (1990) *Nucleic Acids Res.*, **18**, 3998-3998.

65 Fonseca,M.E., Marcellino,L.H., Kitajima,E.W. and Boiteux,L.S. (1994) *J. Gen. Virol.*, **75**, 1447-1449.

66 Rakowski,A.G., Szychowski,J.A., Avena,Z.S. and Semancik,J.S. (1995) *J. Gen. Virol.*, **75**, 3581-3584 (1994)

67 Stasys,R.A., Dry,I.B. and Rezaian,M.A.(1995) *FEBS Lett.*, **358**, 182-184.

68 Koltunow,A.M. and Rezaian,M.A. (1989) *Virology*, 170, 575-578.

69 Koltunow,A.M. and Rezaian,M.A. (1988) *Nucleic Acids Res.*, **16**, 849-864.

70 Rigden,J.E. and Rezaian,A. (1993) *Virology*, 193, 474-477.

71 Hernandez,C., Elena,S.F., Moya,A. and Flores,R. (1992) *J. Gen. Virol.*, **73**, 2503-2507.

72 Symons,R.H. (1981) *Nucleic Acids Res.*, **9**, 6527-6537.

73 Rakowski,A.G. and Symons,R.H.(1989) *Virology*, **173**, 352-356.

74 Pallas,V., Garcia Luque,I., Domingo,E. and Flores,R. (1988) *Nucleic Acids Res.*, **16**, 9864-9864.

75 Semancik J.S., Szychowski J.A. (1994) *J. Gen. Virol.*, **75**, 1543-1549.

76 Rakowski,A.G. and Symons,R.H. (1989) *Virology*, **173**, 352-356.

77 Semancik J.S., Szychowski J.A. (1994) *J. Gen. Virol.*, **75**, 1543-1549.

78 Hernandez,C. and Flores,R. (1992) *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 3711-3715.

79 Beaudry D., Bussière F., Lareau F., Lessard C. and Perreault, J.-P. (1995) *Nucleic Acids Res.*, **23**, 745-752

80 Miller W.A., Hercus T., Waterhouse P.M., Gerlach W.L. (1991) *Virology*, **183**, 711-720.

81 Kaper, J.M.,Tousignant, M.E. and Steger, G. (1988) *Biochem. Biophys. Res. Com.*, **154**, 318-325.

82 Rubino,L., Tousignant,M.E., Steger,G. and Kaper,J.M. (1990) *J. Gen. Virol.*, **71**, 1897-1903.

83 Keese P., Bruening G., Symons R.H. (1983) *FEBS Lett.*, **159**, 185-190.

84 Abouhaidar M.G., Paliwal Y.C. (1988) *J. Gen. Virol.*, **69**, 2369-2373.

85 Buzayan J.M., Gerlach W.L., Bruening G., Keese P., Gould A.R. (1986) *Virology*, **151**, 186-199.

86 Buzayan J.M., McNinch J.S., Schneider I.R., Bruening G. (1987) *Virology*, **160**, 95-99

87 Davies,C., Haseloff,J. and Symons,R.H. (1990) *Virology*, **177**, 216-224.

88 Haseloff,J. and Symons,R.H. (1982) *Nucleic Acids Res.*, **10**, 3681-3691.

89 Makino,S., Chang,M.F., Shieh,C.K., Kamahora,T., Vannier,D.M.,Govindarajan, S. and Lai, M.M. (1987) *Nature*, **329**, 343-346.

90 Epstein L.M., Mahon K.A., Gall J.G. (1986) *J. Cell Biol.*, **103**, 1137-1144.

91 Hernandez,C., Daros,J.A., Elena,S.F., Moya,A. and Flores,R. (1992) *Nucleic Acids Res.*, **20**, 6323-6329.

92 Daros,J.A. and Flores,R. (1995) *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 6856-6860.

93 Koltunow,A.M. and Rezaian,M.A. (1989) *Intervirology*, **30**, 194-201.

94 Elena,S.F., Dopazo,J., Flores,R., Diener,T.O. and Moya,A. (1991) *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 5631-5634.

95 Riesner,D. and Gross,H.J. (1985) *Annu. Rev. Biochem.*, **54**, 531-564.

96 Corpet,F. (1988) *Nucleic Acids Res.*, **16**, 10881-10890.

97 Swofford,D.L. (1992) PAUP: Phylogenetic analysis using parsimony, version 3.0 s (computer program and manual distributed by the Center for Biodiversity, Illinois Natural History Survey, Champaign, IL 61820).

98 Zuker,M. (1989) *Science*, **244**,48-52.

Description du travail

Cette étude a été entreprise dans le but de recenser toutes les séquences se rapportant aux viroïdes et aux ARN apparentés qui étaient disponibles en date du 1er octobre 1995. Nous avons ainsi créer une compilation comprenant les séquences ainsi que toutes les informations pertinentes à chacune d'entre elles (ex: taille, composition, publication, identification des motifs catalytiques, etc.). Nous avons également proposé une nouvelle nomenclature afin de faciliter l'identification et l'utilisation future des séquences de ces ARN. Finalement des études phylogénétiques ont été réalisées afin d'identifier, pour chaque espèce, le variant étant le plus près de la forme ancestrale. La structure secondaire la plus stable de ces variants a été déterminée à l'aide de logiciels informatiques (RNAfold).

Implication dans le travail

Étant impliqué dans un projet sur l'évolution des viroïdes, j'ai proposé la réalisation de cette compilation. J'ai donc été impliqué dans toutes les étapes entourant la mise sur pied et la publication de ce travail. Afin de faciliter la consultation de cette information, nous avons transposé la compilation sur le site WEB http://www.callisto.si.usherb.ca/~jpperra.

CHAPITRE 4:

Compilation des séquences représentant les viroïdes et ARN apparentés

4.2 ARTICLE:

The viroid and viroid-like RNA database

Daniel A. Lafontaine, Patrick Deschênes, **Frédéric Bussière,**

Véronique Poisson et Jean-Pierre Perreault

Département de Biochimie

Université de Sherbrooke

Sherbrooke, Québec, J1H 5N4, Canada

Article publié dans: *Nucleic Acids Research*, Vol. 27, 186-187, 1999.

ABSTRACT

This is an online database in order to facilitate research on viroid, viroid-like RNAs and human hepatitis *delta* virus (vHDV) by presenting a large number of sequences and related data in a comprehensive and user-friendly format (e.g. position of their self-catalytic domains, open reading frame of the vHDV, prediction of the most stable secondary structures, etc.). Most of these RNA species share a common proposed replication pattern known as a DNA-independent rolling circle mechanism. Together, these species form the "brotherhood" of the smallest known auto-replicable RNAs. This online database is available on the World Wide Web (**http://www.callisto.si.usherb.ca/~jpperra**).

INTRODUCTION

The initial version (1996) of this database focused on sequences from viroids, which are small, circular, single-stranded RNA infecting plants, and plant satellite viroid-like RNAs (1). In a subsequent update (1997), we included a section on the human hepatitis *delta* virus (vHDV) (2). The 1998 version was reorganized to be more user-friendly and easier to access the database (3). The four sections (viroids, satellite RNAs, HDV, and others) of the databank are shown in a frame which is always available and allows easy access to any subdivision (see Table I). In addition to the update of the existing database, the 1999 version offers several novel features as described below.

DESCRIPTION

The choice of a section (see Table 1) will lead to the second level which is a summary table for each sequence subdivision. The viroids, satellites RNAs and other RNAs are divided according to the nature of the compilated RNA species. The choice of either a viroid or viroid-like RNA species in a summary level will lead to the third level. Briefly, each species of RNA appearing in the database is listed by its complete name and number of sequence variants. This is followed, for each species, by a complete listing of the sequence variants and their assigned nomenclature. The identification of species variants is based on its usual acronym followed by a number. The procedure for sequence identification and information compilated was presented previously (1). Additional data of each entry include their accession numbers in sequence library file servers, bank loci (when available), number of nucleotides (total and by type), complete publication information, and the sequence in blocks of 10 nucleotides. For several sequences, a table shows structural features, for example the position of the conserved sequences of the self-catalytic domains (hammerhead, hairpin, delta and VS). In addition, a secondary structure prediction of the

most likely ancestral variant of each entry (except for vHDV and VS RNA) was derived using the RNAfold structure prediction package. These predicted structures are appended to the database in connect file format to allow easy manipulation (forth level).

Unlike the viroid section, the vHDV section comprises all available sequences, irrespective of their completeness, since the majority of the partial sequence are informative (i.e. they correspond to the sequence of either the open reading frame or the ribozymes). Both of HDV complete and partial sequences were arbitrarily subdivided in several parts in order to accelerate the display. For the partial vHDV sequence, a "p" preceding the specific number is attributed to partial sequence (2). From 1998 version, sequence alignments of all complete vHDV nucleotide sequences and all amino acid sequences of the vHDV mRNA antigen are available through the summary tables (second level). This database provides an excellent reference point for further phylogenetic and structure-function studies of these RNA species.

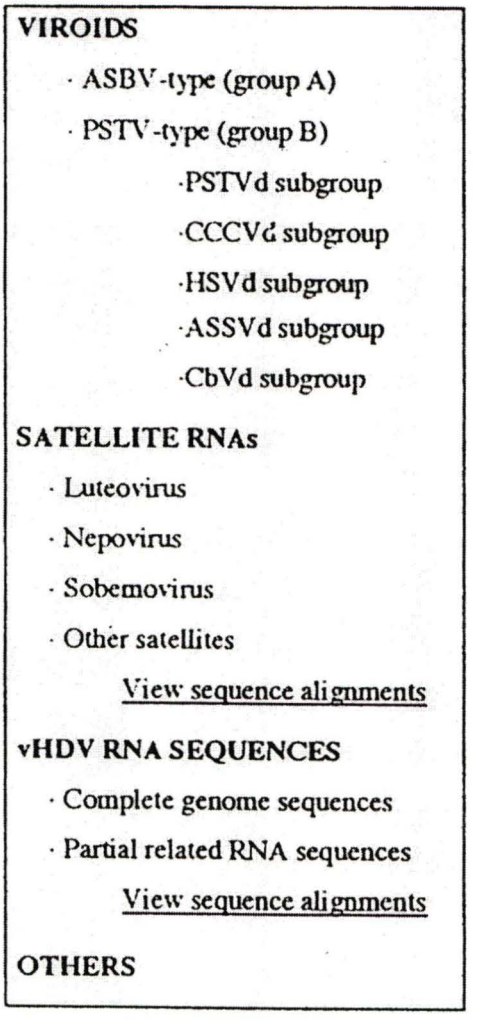


Table 1: Summary of the database sections

1999 VERSION

Today, more than 450 sequences are now part of this database comprising data from 25 viroid species, 9 species of plant satellite viroid-like RNAs, 4 related species of RNA and vHDV (23 complete and 165 partial). The database is now used daily by several researchers, and the user's comments are truly appreciated by the authors. Clearly, the development of the database depends on the input from the members of the research community. The modifications and additions performed in this new version are in response to constructive comments; for example:

i. In agreement with the recent review of Flores and collaborators (4), the group B viroids were clustered in five subgroups according to the sequence composing their central conserved regions (CCR) (see Table 1);

ii. Within the vHDV RNA sequence sections, alignment of all sequence for the self-catalytic domains as well as for the transcriptional promoter are in progress and will be added soon;

iii. Within the satellite RNAs section, the alignment of hammerhead self-catalytic domains is in progress and will be added soon. Subsequently, a phylogenetic reconstruction of the viroids and plant satellite RNA including hammerhead sequence will soon be available; etc.

COMPLETENESS, ACCURACY AND AVAILABILITY OF THE DATA

This databank is available on the World Wide Web browser (e.g. Netscape Navigator) at the URL **http://www.callisto.si.usherb.ca/~jpperra**. Floppy disks (readable on microcomputers operating under MS-DOS or Macintosh) and in hard copies version are also available upon request only for those without electronic access to the database. This database is updated as soon as sequences become available. Users of the viroid and viroid-like RNA database should cite this publication [Lafontaine, D.A., Deschênes, P., Buusière, F., Poisson, V. and Perreault, J.-P. (1999) *Nucleic Acids Res.* **27**] and are encouraged to provide corrections, new information or other information for inclusion in the database via electronic mail (jperre01@courrier.usherb.ca). The authors would appreciate being informed of any omitted sequences or errors in the data set. We will correct any errors.

ACKNOWLEDGEMENTS

This work was supported in part by grants from the Natural Sciences and Engineering Research Council (NSERC), the Medical Research Council (MRC) of Canada, and in part by a scientific team grant from Fonds pour la Formation des Chercheurs et l'Avancement de la Recherche du Québec (FCAR) to J.-P.P. F.B. is recipients of predoctoral fellowships from NSERC while D.L., P.D. and V.P. are recipients of the Fonds de la Recherche en Santé du Québec (FRSQ). J.-P.P. is a MRC schorlar.

REFERENCES

1. Bussière, F., Lafontaine, D. and Perreault, J.-P. (1996) *Nucleic Acids Res.* **24**, 1793-1798.

2. Lafontaine, D., Mercure, S. and Perreault, J.-P. (1997) *Nucleic Acids Res.* **25**, 123-125.

3. Lafontaine, D.A., Mercure, S., Poisson, V. and Perreault, J.-P. (1998) *Nucleic Acids Res.* **26**, 190-191.

4. Flores, R., Di Serio, F. and Hernandez, C. (1997) *Seminars in Virol.* **8**, 65-73.

Description du travail

En 1997 une section traitant du virus de l'hépatite delta humaine (VDH) a été ajoutée (Lafontaine et al., 1997) à la compilation initiale des séquences des viroïdes et ARN apparentés (article 4.1). Le présent article se veut être la mise à jour de la version précédente. Suite aux commentaires reçus, nous avons apporté quelques corrections concernant la classification des viroïdes. De plus l'alignement des séquences des motifs autocatralytiques *delta* et des promoteurs potentiels de l'ARN génomique du VDH a été ajouté. Du côté des viroïdes et ARN satellites de virus de plantes nous avons inclus l'alignement des séquences des motifs autocatalytiques *hammerhead*. Toutes ses modifications seront bientôt disponibles sur le site WEB.

Implication dans le travail

J'ai réalisé le travail entourant l'alignement des motifs autocatalytiques *hammerhead* présents chez les viroïdes du groupe A et chez certains ARN satellites de virus de plantes.

CHAPITRE 5:

Évolution des viroïdes et ARN apparentés.

5.1 ARTICLE:

The hammerhead motifs to attempt solving the evolution of the small, self-cleaving RNA species

Frédéric Bussière et Jean-Pierre Perreault

Département de Biochimie

Université de Sherbrooke

Sherbrooke, Québec, J1H 5N4, Canada

Article en rédaction.

Introduction

Viroids, viroid-like plant satellite RNAs and the hepatitis *delta* virus (HDV) are small (246-1700 bases), single-stranded, circular RNA pathogens (Table 1) (reviewed by Flores et al., 1997; Symons, 1997). At the exception of HDV, which it is retreived in mammals and codes for an antigen, all these RNAs are found in plants and do not encode any polypeptide. Since these RNA species share numerous molecular features, they have been proposed to form a brotherhood (Branch et al., 1990). Based on a phylogenetic reconstruction, Elena et al. (1991) suggested that the automously replicating viroids and viroid-like satellite RNAs, which require a helper virus for their replication, form two distinct monophyletic groups. ASBVd, the only self-cleaving viroid known at that time was proposed to form a new cluster (group A viroid) representing an evolutionary link between the two others. However, recent sequence analysis of these species was not able to find any basic similarity between viroids, satellite RNAs and HDV (i.e. meaning that these sequences cannot be aligned in a statistically significant way; Jenkins et al., 1999; Bussière, 1999). The sequence homologies were restricted between the 23 members of the group B viroids that also share a genomic organization including five structural domains and a conserved central region (CCR). More importantly, sequence data alone cannot support the claim of a common evolutionary origin for the RNA species of this brotherhood. Therefore, we decided to review the evolutionnal relationships of these RNA species from another angle.

Rolling-circle replication as an informative biological mechanism and self-cleaving hammerhad RNA motif as a conserved feature

These RNA species replicates through a DNA-independent rolling-circle mechanism (Branch & Robertson, 1983). This mechanism involves the synthesis of multimeric strands and their subsequent cleavage into monomeric fragments, which are then circularized to

TABLE 1. Viroid and viroid-like RNAs.

Species name	abbreviation	RNA type	sub-classification	catalytic motifs [a] (+) / (-) [b]	replication mode [c]
Potato spindle tuber viroid [d]	PSTVd	viroid	group B	-/-	asym.
Avocado sunblotch viroid	ASBVd	viroid	group A	hh / hh	sym.
Peach latent mosaic viroid	PLMVd	viroid	group A	hh / hh	sym.
Chrysanthemun chlorotic mottle viroid	CChMVd	viroid	group A	hh / hh	sym.
Cherry small circular viroid-like RNA 1	cscRNA1	?	?	hh / hh	sym.
Barley yellow dwarf virus satellite RNA	vBYDV	satellite	luteovirus	hh / hh	sym.
Lucerne transient streak virus satellite RNA	vLTSV	satellite	nepovirus	hh / hh	sym.
Arabic mosaic virus satellite RNA	sARMV	satellite	nepovirus	hh / hp	sym.
Tobacco ringspot virus satellite RNA	sTobRV	satellite	nepovirus	hh / hp	sym.
Chicory yellow mottle virus satellite RNA (S1)	sCYMV	satellite	nepovirus	hh / hp	sym.
Rice mottle virus satellite RNA	sRYMV	satellite	nepovirus	hh / -	asym.
Velvet tobacco mottle citrus satellite RNA	vVTMoV	satellite	sobemovirus	hh / -	asym.
Solanum nodiflorum mottle virus satellite RNA	vSNMV	satellite	sobemovirus	hh / -	asym.
Subterranean clover mottle virus satellite RNA	vSCMoV	satellite	sobemovirus	hh / -	asym.
Hepatitis delta virus	vHDV	satellite	hepadnavirus	delta/delta	sym.
Carnation stunt associated viroid	rCarSV	retroviroid	---	hh / hh	n.a.

[a] hh, hp indicate hammerhead and hairpin motifs.

[b] plus and minus strands.

[c] proposed replication mode is either symmetric (sym.) or asymmetric (asym.). n.a. indicates not applicable.

[d] PSTVd is an example among the 26 viroids of group B.

produce the progeny (Fig. 1). Regarding whether or not the minus circular strands are produced and used as templates, this mechanism as two modes, i.e. symmetric or asymmetric. The replicational mode is one of the hallmark to differentiate in which of the two groups (i.e. A or B) belong a viroid (Table 1) (Bussière et al., 1999). Group B viroids, such as PSTVd, are believed to replicate according to the asymmetric mode involving an host endonuclease activity to process concatameric strands of plus polarity. Instead viroids of group A, like ASBVd, are suggested to replicate via the symmetric mode and share self-cleaving hammerhead motifs mediating the processing of the strands of both plus (+) and minus (-) polarities. Similarly viroid-like satellite RNAs are proposed to replicate according to one of this two modes depending either the strand of (+) polarity alone or both strands possess self-cleaving RNA motifs (Table 1). In theirs cases, it is always a self-catalytic RNA motif that mediate the cleavage of the concatameric strands. The 3 group A viroids and the 11 plant satellite RNAs possess a hammerhead motif at least on their plus strand ((+) hammerhead), and 7 of them on strands of both polarities (Table 1). The 7 later species as well as the 3 with a hairpin motif on the strand of minus polarity are believed to replicate through a symmetric rolling-circle mechanism. In contrast, the 4 sobemovirus satellite RNAs, having only a (+) hammerhead, are proposed to replicate through the asymmetric mode. Finally, HDV possess self-catalytic delta motif on both polarities strands and thus replicate according to the symmetric mode.

At the exception of HDV that has been proposed to evolved from a mRNA captured by a primitive viroid-like RNA (Branch et al., 1990; Brazas & Ganem, 1996), the hammerhead motif appears the only "homologous character" shared by all self-cleaving circular RNAs (i.e. group A viroid and satellite RNAs). Hammerhead motifs were also found in cellular transcripts, which have been proposed to be retroposons, from repetitive DNA of salamanders (rNS2T) and schistosomes (rSmaa) (Epstein & Gall, 1987; Febeyre et al., 1998). These transcripts probably do not involved strands of minus polarity as an intermediate for their retroposition, and thus possess only a (+) hammerhead (Table 1).

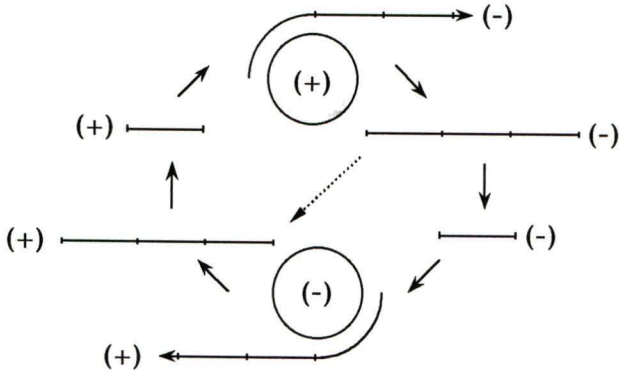


FIGURE 1. Rolling circle replication.

The rolling-circle replication. (+) and (-) indicate the strand polarities and the dotted central arrow is for the asymmetric mode.

In most of the self-cleaving species (11/16), (+) hammerhead sequences have the same linear organization in the genome (Fig. 2A and D). Two short hairpins ranging between 12 and 20 nt form stems I and II of the hammerhead. The remaining nucleotides of the RNA genome, refered here as "extracatalytic" RNA, are located in the stem III region, although it could be theorically found in either of the two other stems. These (+) hammerhead possess stable stem III and consequently are proposed to self-cleave through single hammerhead structures (reviewed by Symons, 1992). This conserved genomic organization of the hammerhead motif suggest a common evolutionary origin. The (+) hammerhead of vBYDV differ slightly since it contains an insertion of 54 nt in the stem II (Miller et al., 1991). The 5 other RNA species do not harbor the aforementionned hammerhead organization, their "extracatalytic" RNA sequences are rather located in the stem I region (Fig. 2B and D). In addition, an important portion of the ASBVd and sRYMV genome is also located in the stem II region leading to the segragation of the catalytic core in two domains (Fig. 2C and D). In these five species, the (+) hammerhead stem III is composed by only a few nucleotides, and thus is relatively unstable and self-cleavage was proposed to occur via double hammerhead structures involving longer-than-unit RNA molecules to allow stabilization of the catalytic core (Symons, 1992).

Most of the (-) hammerhead structure share the same organization that their (+) counterpart (Fig. 2D). The exceptions to this observation are with 3 species. vBYDV did not have an insertion in hammerhead stem II while CChMVd has one in this stem. Finally, rCarSV has the genomic organization of most of the plus strand RNA species with all "extracatalytic sequence within its stem III, while its (+) strand is of the type with the "extracatalytic sequence in stem I. Regardless these exceptions, the organizational similarities of the hammerhead motifs within strands of both polarities should be informative. One potential consequence is that the regions spanning both hammerheads can fold into long rod-like structures. This typical arrangement is found in the most stable secondary structure predictions of several species, like cscRNA1 and PLMVd (Di Serio et

al., 1997; Beaudry et al., 1995). The rod-like structure formed by the superposed hammerhead sequences has the potential to prevent the subsequent folding into the catalytic active structures inhibiting self-cleavage. The two hammerheads of vBYDV can not be superposed because the (+) hammerhead possess an insertion in the stem II (54 nt). In this case, self-cleavage inhibition has been suggested to result from formation of a pseudoknot involving bases of the insertion and stem I (Miller et al., 1991). It seems that self-cleavage inhibition maybe a mechanism widely used to stabilize the circular conformers that are the templates in the rolling-circle mechanism. Alternatively, these rod-like structures may contribute positively to the extra- (e.g. during transmission) and intracellular stability of these RNA species. More likely, a selective pressure had favored the apparition of the superposed hammerhead sequences.

The superposition of the hammerhead sequences may allows to give some clues about the molecular evolution of these RNA species. In order to reach superposed hammerheads, one possibility is that a (+) hammerhead had served as template for the synthesis of its counterpart either by RNA recombinason (template switching; Forster & Symons, 1987) during replication or progressively by successive mutations. Regardless of the mechanism, the first (-) hammerhead would have to be synthesized by copying the (+) hammerhead.

Hammerhead motif is formed by a catalytic pocket, for which the bases are highly conserved, surrounding by three stems shown to be quite independent regarding both their sequence and size for efficient self-cleavage *in vitro* (Tuschl & Eckstein, 1993; Clouet-d'Orval & Uhlenbeck, 1997). Considering that viroids and related species are RNA genomes and consequently mutated rapidly, we could expected that the stems of their hammerhead to be not conserved. Contrastly, the observation of the sequence conservation forming the stem would strongly support the notion that the hammerhead is a common character within these species. Panel D of figure 4 show the sequence alignment of all hammerhead motifs while panel E illustrates the secondary structure and the conserved

nucleotides of the hammerhead motifs. Despite the lack of important nucleotides in the stems for efficient hammerhead self-cleavage *in vitro,* we noticed an unexpected conservation of the sequence of the hairpin forming the stem II in the plus strands of natural isolates (Fig. 2D-E). A purine-rich loop of 4 to 6 nt and a 4 to 5 bp stem which shows a strong consensus, i.e. 5'-$G_{10.1}$KCC-(purine rich)-GGM$C_{11.1}$-3' (where K and M indicate G or U and C or A, respectively). The $G_{10.1}$-$C_{11.1}$ base pair that was shown *in vitro* to be crucial, is retreived in all RNA sequences. It was demonstrated *in vitro* that the size of the stems I and II can has deleterious effect when longer than 5 and 4 bp, respectively (Tuschl & Eckstein, 1993, Clouet-d'Orval & Uhlenbeck, 1997); a long stem I (> 5 pb) does not inhibit the ribozyme only if stem II is shorter than 5 bp, *vice et versa* the presence of a long stem II does not have inhibitory effect only in presence of a short stem I. Analysis of the stems I and II length shows that most RNA species avoid the stem lenght inhibition. In contrast to stem II, for which conservation appear to span through all species, the composition of stem I show sequence similarity only between those that are closely related, like vSCMoV, vSNMV and vTMoV. Interestingly though to be a satellite RNA, cscRNA1 possesses a stem I highly similar to known sobemovirus associated RNA, as vSNMV. Therefore, not only the (+) hammerhead motifs and their organizations are conserved, but also their structural features (i.e. primary and secondary structures). Similarly, (-) hammerheads have the aforementionned structural organisation including the same hairpin II consensus while their stem I is less conserved.

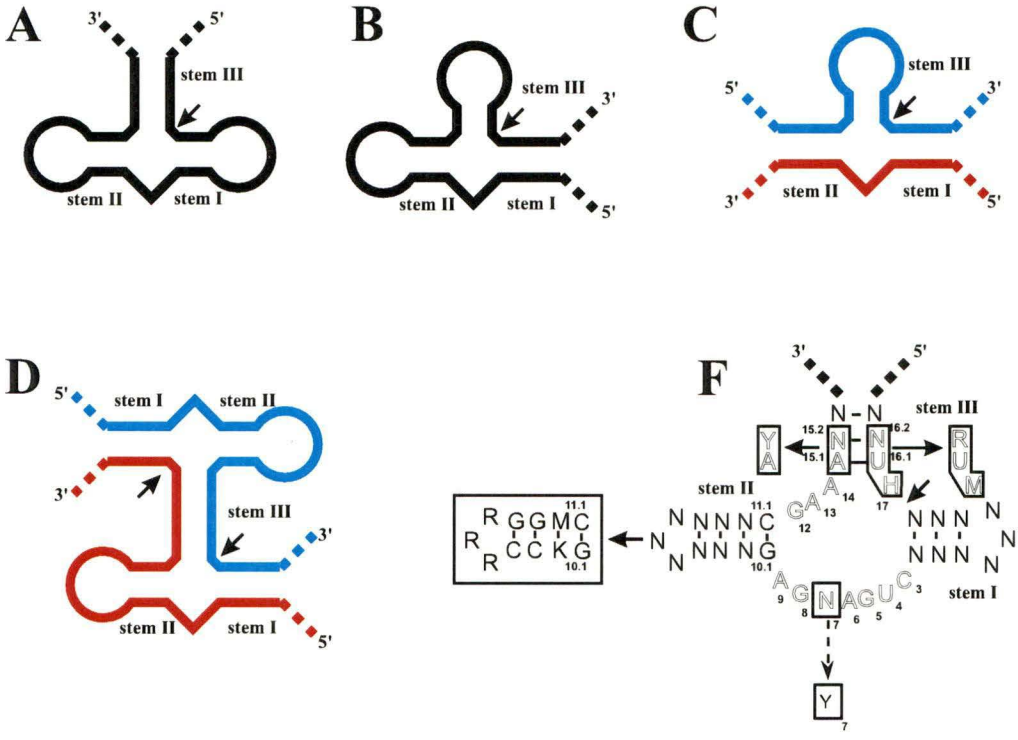

A
3'
5'
stem III
stem II
stem I
B
stem III
3'
stem II
stem I
5'
C
stem III
5'
3'
3'
stem II
stem I
5'
D
5'
stem I
stem II
3'
stem III
3'
stem II
stem I
5'
F
3'
5'
N-N
stem III
stem II
stem I
E
NUH
stem-loop I
17-20nt
Catalytic
core
stem-loop II
12-20nt
GAAAN
(+)
vBYDV
sARMV
sCYMV
sTobRV
vSCMoV
vSNMV
vVTMoV
vLTSV
cscRNA1
PLMVd
CChMVd
rCarSV
ASBVd
sRYMV
stem III
stem I
NUH
CONSENSUS
(-)
vBYDV
vLTSV
cscRNA1
PLMVd
CChMVd
rCarSV
ASBVd
stem III
stem I
NUH

Figure 2 (page précédente). Conservation of the hammerhead structures and sequences and model genomic organization. A and B Hammerheads of species with the "extracatalytic" sequence in the stem III and II, respectively. **C** ASBVd-type hammerhead in which the involves sequences are non-contiguous. While the hammerhead in process through a single structure, those in B and C required formation of a double structure, illustrated at rigth. **D** Organization and sequences of the hammerhead found on both the plus (+) and minus (-) polarities RNAs. The size of stem-loops and the positions of the loops (underlined nucleotides) are shown. - are gaps while insertion size are between parenthesis. Boxed nucleotides in stems are not involved in Watson-Crick interactions. Arrows represent the position of extracatalytic sequence. When several variants were known a consensus sequence was used (R= A or G; Y= U or C; W= A or U; K= G or U; M= C or A). The written consensus of (+) hammerhead stem-loop II represents the sequence found in at least 11 out of 13 considered RNAs. Sequence where retreived from the viroid and viroid-like database (http://www.callisto. si.usherb.ca/~jpperra; Lafontaine et al., 1999). **E** Nucleotide sequence and secondary structure of the hammerhead self-catalytic motif. The three conserved domains are outshadowed. Conserved positions in RNA species are boxed. N: A, G, C or U; I: inosine; H: A, C or U. The arrow points the cleavage site.

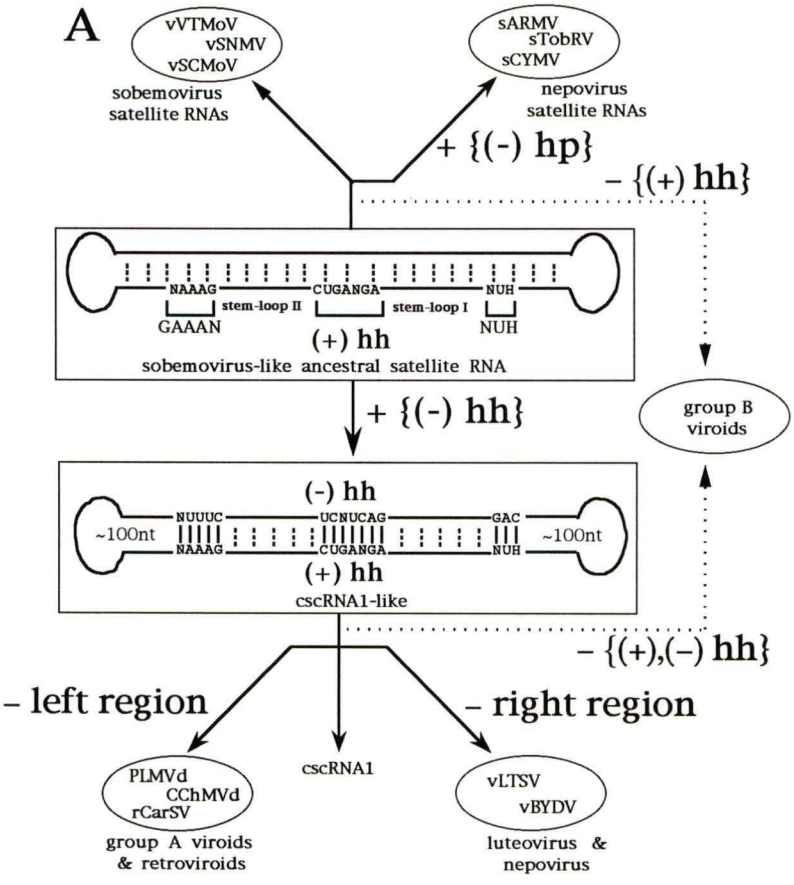
A potential evolutional reconstruction

The conservation of the genomic organization and sequence of both (+) and (-) hammerhead motifs suggest that it is a molecular features that have been conserved during the evolution of the group A and plant viroid-like satellite RNAs. Using these informations, we attempted to define an potential evolutional reconstruction of these RNA species (Fig. 3A).

The ancestral RNA species, would have possessed a hammerhead motif solely on its (+) strand, like sobemovirus satellite RNAs. Then, selective pressures (see above) triggered the apparition of a self-cleaving motif for the processing of (-) multimeric strands in two lineages: a hairpin self-catalytic motif for the lineage leading to the nepoviruses satellite RNAs (except vLTSV), and a (-) hammerhead for the RNA species possessing two of these motifs. Regarding the position of the hammerhead sequences as compared to the "extracatalytic" ones, the latest species could be classified in three distinct clusters (see Fig. 2D and 3A). The hammerheads form either: i. the left-hand region, like PLMVd; ii. the right-hand region, like vLTSV; or, iii. the central region and thus "extracatalytic" sequences are distributed equally in both sides, like cscRNA1. Considering the surprising similarity between the stem I region of cscRNA1 and the sobemovirus satellite RNA, it is tempting to speculate that a cscRNA1-like RNA would be the direct ancestor of the three clusters. It is noteworthy that stem I conservation could be the result of close relationship. By a deletion event in either the left or the right terminal region, this ancestor have given rise to either the group A viroids (e.g. PLMVd), or the cluster formed by vBYDV and vLTSV. Simple mechanism such as spontaneous deletion in RNA sequences have been shown to occur on a biologically meaningful time scale (reviewed in Chetverin, 1999; Nagy & Simon, 1997). The different organization of the (+) hammerhead of rCarSV most probably arose after the divergence of the group A viroids cluster (Hernandez et al., 1992). Moreover, the transition of rCarSV to a DNA element (Daros & Flores, 1995), may have contribute to selective pressures that differ from viroids which use solely DNA-independent replication.

This reconstruction implies that hammerhead containing viroids form a monophyletic cluster which evolved from a satellite RNA. It requires minimal number of evolutionnal events, and thus appears parcimonious. More importantly, this reconstruction either take into account or respect all known biological characteristics of these RNAs.

Group B viroids do not share any nucleotide or structural similarity with either group A viroids or other hammerhead containing species, therefore, their evolutional relationship remains obscure. If they share a common origin with the other species of the brotherhood, they most likely arose from either the sobemovirus-like ancestral satellite RNA or a direct ancestor of group A viroids (see Fig. 3A). The latter hypothesis implies the lost of both the (+) and (-) hammerheads and the convergent evolution of viroid features in two independent lineages; one leading to group A and another leading to group B. According to this hypothesis, viroids would form a monophyletic group, as suggested previously (Elena et al., 1991). The alternative hypothesis, i.e. from a sobemovirus-like satellite RNA, requires only the disappearance of the (+) hammerheads. However, this hypothesis implies that viroids form a polyphyletic groups. Clearly, viroids known yet do not allow to solve this issue.


A
vVTMoV
vSNMV
vSCMoV
sobemovirus
satellite RNAs
sARMV
sTobRV
sCYMV
nepovirus
satellite RNAs
+ {(-) hp}
– {(+) hh}
NAAAG
CUGANGA
NUH
stem-loop II
stem-loop I
GAAAN
NUH
(+) hh
sobemovirus-like ancestral satellite RNA
+ {(-) hh}
group B
viroids
(-) hh
NUUUC
UCNUCAG
GAC
~100nt
~100nt
NAAAG
CUGANGA
NUH
(+) hh
cscRNA1-like
– {(+),(–) hh}
– left region
– right region
PLMVd
CChMVd
rCarSV
group A viroids
& retroviroids
cscRNA1
vLTSV
vBYDV
luteovirus &
nepovirus

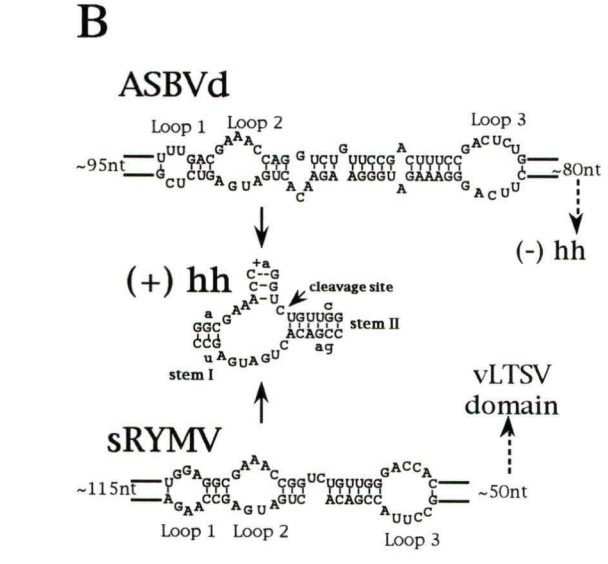

B
ASBVd
Loop 1
Loop 2
Loop 3
~95nt
~80nt
(-) hh
(+) hh
cleavage site
stem II
stem I
vLTSV
domain
sRYMV
~115nt
~50nt
Loop 1
Loop 2
Loop 3

Figure 3. Hypothetical reconstruction of viroid and viroid-like RNAs evolution. A General scheme (see text for details) (page précédente). Plus (+) and minus (-) between parenthesis identified the strands on which a haamerhead (hh) or hairpin (hp) is modified. Other plus and minus symbols indicate respectively the addition or substraction of sequences. **B** Comparison of the central regions of ASBVd and sRYMV. The right terminal domain of ASBVd include the (-) hammerhead sequences (hh (-)) while this region in sRYMV is almost identical to a portion of the vLTSV. In the middle is the (+) hammerhead structure form by this RNA species. The uppercase sequence corresponds to sRYMV while lowercase letter indicate comparative ASBVd mutations.

Species evolving from group A viroids

ASBVd and sRYMV are two species for which the evolutional relationships are not simple to attempt. For example, ASBVd is the only species with non-contiguous hammerhead structures (divided in two domain). However, since it is a group A viroid, parcimony would favor its evolution from the same lineage that PLMVd and CChMVd. Furthermore, we observed that sRYMV, a recently discovered nepovirus satellite RNA (Collins et al., 1998), has an ASBVd-type hammerhead structures (Fig. 3B). Both the ASBVd and sRYMV species are of the type with "extracatalytic" sequences both side of the (+) hammerhead structure contained in the central region. Moreover, both the architecture and sequence of the central region in their most stable secondary structure, appears closely related (Fig. 3B). However, it is unknow whether or these similarities have functional significance. In addition, the sequence of the 50-nt long right domain of sRYMV has been reported to be nearly identical to a vLTSV region (Collins et al., 1998). Taking together these observations, we propose that sRYMV is a recombinant species between a ASBVd-like RNA and vLTSV.

Is it rational that viroids evolved from satellite RNAs

No characterized RNA species has been identified yet to serve as an appropriate outgroup in order to infer the origin of viroids and viroid-like satellite RNAs as well as to establish the direction of their evolution. In counterpart, the biology of these RNA species is instructive. According to our reconstitution, we propose that a sobemovirus-like RNA diverge to give rise to today's viroids. Satellite RNAs use proteins encoded by both their helper virus and the host to support their life cycle including their replication (reviewed in Roossinck et al., 1992). For example, a given satellite RNA may require the viral encoded movement protein

to allow its recognition by the plasmodesmata structural features and its further translocation into the adjacent cells. Since the satellite RNA already interacts with plasmodesmata components, it is likely to evolve adaptative changes to completely usurp this normal cellular RNA trafficking machinery. Thus, the viral protein is no longer needed and the satellite RNA become independent from its helper virus. Satellite RNAs could benefit from such progression adaptations instead of major evolutionnary steps. The fact that host encoded proteins play a role in most steps, or at least some steps, of their life cycle make these RNAs susceptible to acquire their indepency from the helper virus to survive and propagate into the cellular environment. In other words, they become viroids. Most of the time satellite RNAs are in contact with their host, hence have the opportunity to evolve the new interactions that would promote their selfish viability in the infected cells (Roossinck et al., 1992). Unlike satellite RNAs, viroids are less likely to encounter a given virus and to adapt itself quickly to be encapsidated and eventually replicated by the viral machinery. The capability of some viroids to be encapsidated in viral particules or to be transmitted by aphid vectors is probably a remnant of their satellite RNA origin (Desvignes, 1986; Querci et al., 1997).

This reconstruction does not rule-out the possibility that viroids were the ancestor of satellite RNAs. However, we think it would require the involvement of least propable major evolutional step instead of only molecular adaptation. Regardless the evolutional scheme, the plasticity of these RNA species (i.e. high mutation rate) has certainly contributed positively to their evolution, more specifically to their adaptation. In addition, the movement into different cellular compartments (i.e. group A in chloroplast and group B in nucleus; Bussière et al, 1999), had been of primary importance for the evolution of these RNA genomes.

The escaped self-cleaving transcripts

The most puzzling question is the putative relationship between these RNA pathogens and the hammerhead self-cleaving transcripts from satellite DNA found in salamanders (rNS2T) and schistososmes (rSmaa) (Epstein & Gall, 1987; Febeyre et al., 1998). As proposed by Ferbeyre et al., the rNS2T and rSmaa elements could be homologous retrotranposons. This proposal receives additionnal support from the recent characterization of hammerhead sequences in repetitive DNA elements of a wide range of unrelated species, and were proposed to evolve from a common ancestor (Laferrière et al., 1994). Thus, although they have not yet been found in plants, such *cis*-acting ribozymes containing retrotranposons are likely to be found in their genome. No information is currently available to indicate the direction of the evolution between viroid and satellite RNA to these transcripts. The most simple scenario is that a viroid or a satellite RNA was integrated in a genome, and then lost most of its viroid-like characters including the DNA-independent replication. However, the inverse scenario that involves that such a transcript was at the origine of the first member of the brotherhood, is also possible. In this case, the transcript would have escape the DNA replication and become a viroid or a satellite RNA. Since they did not need further transcription from DNA regions, these elements acquired the capability to be horizontally transferred into other "permissive" plants. The need for self-cleavage reaction in RNA based rolling-circle replication would have led to the conservation of the (+) hammerhead motif.

In summary, the important divergence of genome sequences among viroids and related RNAs impeded classical phylogenetic reconstructions and hence the study of their putative evolutionnal relationships. A more detailed analysis of the hammerhead sequences strongly support a common origin of viroids and plant viroid-like satellite RNAs, allowing us to elaborate a new evolutionnary scenario. Obviously, discovery of novel species, would permit both to complete this evolutional story. Clearly, viroids and related species, which

have been proposed as model "relic" from a precellular RNA world, remain intriguant RNA organisms provide many evolutional and biological issues to address in the coming years.

ACKNOWLEDGEMENTS

The authors would like to thank their collegues B. Cousineau and R. Cedergren for critical discussions. This work was supported in part by grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada to JPP. FB was recipient of a predoctoral fellowship from the NSERC. JPP is a MRC scholar.

REFERENCES

Beaudry D, Busiere F, Lareau F, Lessard C, Perreault JP. 1995. The RNA of both polarities of the peach latent mosaic viroid self- cleaves *in vitro* solely by single hammerhead structures. *Nucleic Acids Res 23*:745-752.

Branch AD, Robertson HD. 1983. A replication cycle for viroids and other small infectious RNAs. *Science 223*:450-455.

Branch AD, Levine BJ, Robertson HD. 1990. The brotherhood of circular RNA pathogens: viroids, circular satellite, and the delta agent. *Semin Virol 1*:143-152.

Brazas R, Ganem D. 1996. A cellular homolog of hepatitis delta antigen: implications for viral replication and evolution. *Science 274*:90-94.

Bussière F, Lafontaine D, Perreault JP. 1996. Compilation and analysis of viroid and viroid-like RNA sequences. *Nucleic Acids Res 24*:1793-1798.

Bussière F. 1999. Le viroïde de la mosaïque latente du pêcher (PLMVd): un modèle pour mieux comprendre l'évolution et la biologie des viroïdes du groupe A. Ph. D. thesis. Université de Sherbrooke, Qc, Canada.

Bussière F, Lehoux J, Thompson DA, Skrzeczkowski LJ, Perreault JP. 1999. Subcellular localization and the rolling circle replication of peach latent mosaic viroid: hallmarks of group A viroids. *J Virol 73*:6353-6360.

Clouet-d'Orval B, Uhlenbeck OC. 1997. Hammerhead ribozymes with a faster cleavage rate. *Biochemistry 36*:9087-9092.

Chetverina AB. 1999. The puzzle of RNA recombination. *FEBS lett 450*:1-5.

Collins RF, Gellatly DL, Sehgal OP, Abouhaidar MG. 1998. Self-cleaving circular RNA associated with rice mottle virus is the smallest viroid-like RNA. *Virology 241*:269-275.

Daros JA, Flores R. 1995. Identification of a retroviroid-like element from plants. *Proc Natl Acad Sci USA 92*:6856-6860.

Desvignes JC. 1986. Peach latent mosaic and its relation to peach mosaic and peach yellow mosaic virus disease. *Acta Horticul 235*:325-332.

Di Serio F, Daros JA, Ragozzino A, Flores R. 1997. A 451-nucleotide circular RNA from cherry with hammerhead ribozymes in its strands of both polarities. *J Virol 71*:6603-6610.

Elena SF, Dopazo J, Flores R, Diener TO, Moya A. 1991. Phylogeny of viroids, viroidlike satellite RNAs, and the viroidlike domain of hepatitis *delta* virus RNA. *Proc Natl Acad Sci USA 88*:5631-5634.

Epstein LM, Gall JG. 1987. Self-cleaving transcripts of satellite DNA from the newt. *Cell 48*: 535-543.

Ferbeyre G, Smith JM, Cedergren R. 1998. Schistosome satellite DNA encodes active hammerhead ribozymes. *Mol Cell Biol 18*:3880-3888.

Flores R, Serio F, Hernandez C. 1997. Viroids: The noncoding genomes. *Seminars Virol 8*:65-73.

Forster AC, Symons RH. 1987. Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. *Cell 49*:211-220.

Hernandez C, Daros JA, Elena SF, Moya A, Flores R. 1992. The strands of both polarities of a small circular RNA from carnation self-cleave *in vitro* through alternative double- and single-hammerhead structures. *Nucleic Acids Res 20*:6323-6329.

Jenkins GM, Woelk CH, Rambaut A, Holmes E. 1999. Testing the extend of sequence similarity among viroids, satellite RNAs and hepatitis delta virus. *J Mol Evol*:in press.

Laferriere A, Gautheret D, Cedergren R. 1994. An RNA pattern matching program with enhanced performance and portability. *Comput Appl Biosci 10*:211-212.

Lafontaine DA, Deschênes P, Bussière F, Poisson V, Perreault JP. 1999. The viroid and viroid-like RNA database. *Nucleic Acids Res. 27*:186-187.

Miller WA, Silver SL. 1991. Alternative tertiary structure attenuates self-cleavage of the ribozyme in the satellite RNA of barley yellow dwarf virus. *Nucleic Acids Res 19*:5313-5320.

Nagy PD, Simon, AE. 1997. New insights into the mechanism of RNA recombination. *Virology 235*:1-9.

Querci M, Owens RA, Bartolini I, Lazarte V, Salazar LF. 1997. Evidence for heterologous encapsidation of potato spindle tuber viroid in particles of potato leafroll virus. *J Gen Virol 78*:1207-1211.

Roossinck MJ, Sleat D, Palukaitis P. 1992. Satellite RNAs of plant viruses: structures and biological effects. *Microbiol Reviews 56*:265-279.

Symons RH. 1992. Small catalytic RNAs. *Ann. Rev. Biochem. 61*:641-671.

Symons RH. 1997. Plant pathogenic RNAs and RNA catalysis. *Nucleic Acids Res 25*:2683-2689.

Tuschl T, Eckstein F. (1993) Hammerhead ribozymes: importance of stem-loop II for activity. *Proc Natl Acad Sci USA 90*:6991-6994.

Description du travail

L'objectif principal de ce travail était l'étude des relations évolutives au sein des viroïdes et des ARN apparentés. La grande divergence au niveau des séquences de ces ARN rend impossible l'utilisation des approches phylogénétiques conventionnelles. Cette situation est encore plus problématique lorsque l'on tente de déterminer les liens évolutifs entre les ARN satellites et les viroïdes. Le motif autocatalytique *hammerhead* semble être le seul domaine conservé chez les viroïdes du groupe A et les ARN satellites. Une étude plus approfondie du *hammerhead* supporte une origine évolutive commune pour ces ARN. Ce motif catalytique a donc été utilisé comme caractère homologue afin de reconstruire leur histoire évolutive. Le modèle proposé suggère, entre autres, qu'un ARN satellite, ayant su s'adapter pour devenir indépendant de son virus associé, aurait donner naissance aux viroïdes actuels. La rédaction de l'article est complétée et il reste principalement un travail d'édition du texte. Le manuscrit sera donc soumis sous peu à la revue *RNA*.

Implication dans le travail

J'ai réalisé la totalité du travail.

DISCUSSION

1. LA BIOLOGIE MOLÉCULAIRE DE PLMVD ET DES ARN APPARENTÉS

1.1 L'autocoupure

Le motif autocatalytique *hammerhead* est présent chez les viroïdes du groupe A et chez certains ARN satellites de virus de plantes (voir section 5.1 des Résultats). Tous ces ARN possèdent un *hammerhead* qui permet l'autocoupure des multimères (+) produits durant la réplication en cercle roulant. Les ARN ayant un mécanisme en cercle roulant symmétrique possèdent également un *hammerhead* ou un motif en épingle à cheveux (*hairpin*) pour assurer la coupure des brins (-). À l'intérieur du génome, le *hammerhead* se retrouve dissimulé dans un long ARN riche en structures secondaires. La réaction d'autocoupure dépend donc de la capacité de former une structure catalytique active en dépit des nombreuses structures alternatives. Chez PLMVd, la région qui compose le *hammerhead* se retrouve dans une longue hélice très stable qui empêche l'adoption des structures catalytiques. Cependant, durant la transcription, la synthèse du *hammerhead* précède celle de la région inhibitrice complémentaire, permettant l'adoption transitoire d'une structure active stable (fig. 6). Il en résulte une réaction d'autocoupure intramoléculaire très efficace. Éventuellement, la longue hélice d'ARN est recréée, empêchant ainsi une nouvelle réaction d'autocoupure (section 1.1 des Résultats). Cette inhibition pourrait être importante afin de favoriser l'accumulation des formes circulaires qui seraient linéarisées par une réaction d'autocoupure. L'organisation consécutive des trois blocs conservés formant le *hammerhead* est probablement cruciale pour favoriser la formation transitoire des tiges I, II

et III. La stabilité de la tige III assure une réaction en simple *hammerhead* chez PLMVd (section 1.1 des Résultats). Chez la majorité des viroïdes et ARN satellites, l'organisation des *hammerhead*, ainsi que leur positionnement à l'intérieur de longues hélices, rapelle les observations faites chez PLMVd. Les mécanismes de régulation de l'autocoupure proposés pour PLMVd semblent donc pouvoir s'appliquer à la majorité d'entre eux.

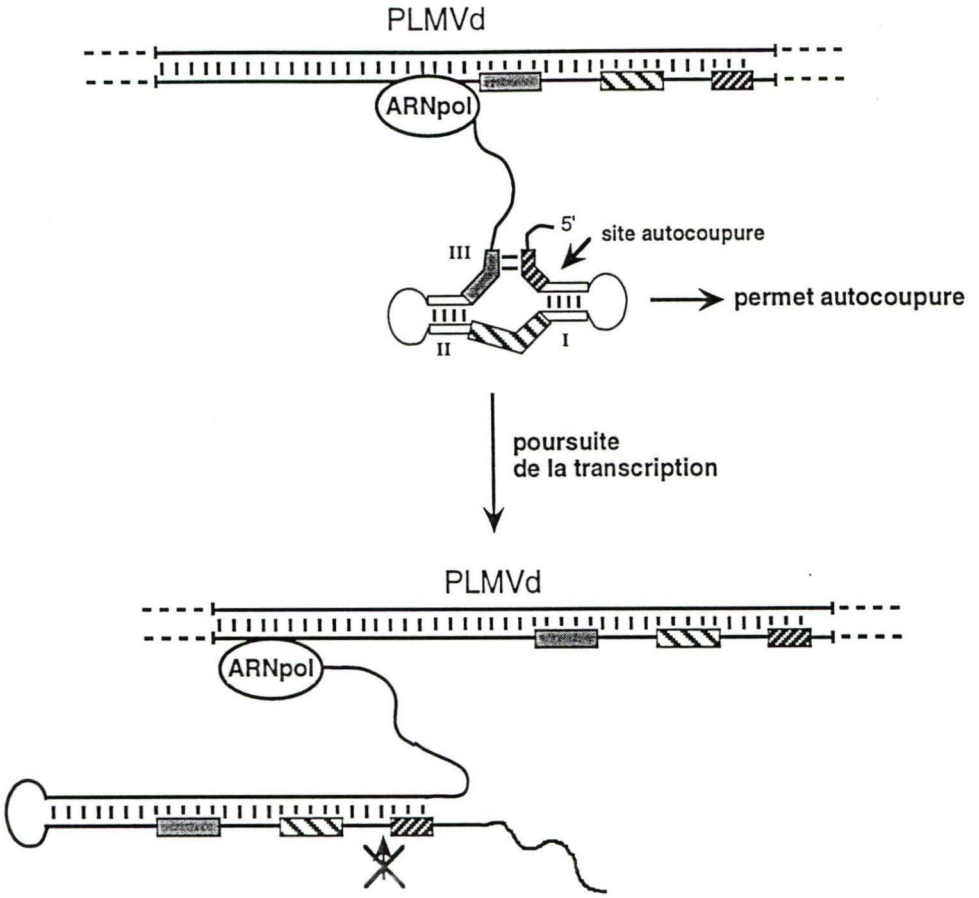


Figure 6. Formation transitoire de la forme active du *hammerhead* durant la transcription, chez PLMVd. L'ARN polymérase (ARNpol) transcrit la région qui contient le *hammerhead*, permettant ainsi la réaction d'autocoupure. La prochaine portion à être synthétisée mène à la formation d'une longue hélice d'ARN qui empêche l'adoption de la structure catalytique active. L'autocoupure des transcrits n'ayant pu être préalablement coupés est ainsi inhibée. La figure schématise la partie double brin du génome de PLMVd qui contient les séquences formant les *hammerhead* des deux polarités (voir section 1.1 des Résultats).

L'inhibition post-transcriptionnelle de l'autocoupure est illustrée par un mécanisme différent chez vBYDV. Cet ARN satellite possède un *hammerhead* (+) dont l'organisation s'apparente à celle de PLMVd, c'est-à-dire que les trois blocs conservés sont consécutifs et séparés par de courtes tige/boucles formant les tiges I et II (section 5.1 des Résultats). Les prédictions de structures secondaires par ordinateur montrent que le motif catalytique est situé dans une région instable capable d'adopter de nombreuses structures alternatives. Dans un tel contexte, une structure en simple *hammerhead* pourrait être aisément formée. La forme native du génome de vBYDV ne serait donc pas en mesure d'empêcher l'activation de l'autocoupure. Cependant, chez vBYDV la tige/boucle II diffère puisqu'elle possède une insertion d'environ 50 nt qui interagit avec la boucle de la tige I pour former une structure tertiaire appelée pseudo-noeud. Celle-ci inhibe la réaction d'autocoupure en simple *hammerhead* (fig. 7) (Miller et Silver, 1991). Durant la réplication, une structure en double *hammerhead* empêcherait la formation du pseudo-noeud et par conséquent l'inhibition de la réaction d'autocoupure (communication personnelle de S.I. Song). Cette interaction en *trans* entre deux *hammerhead* est probablement favorisée par l'instabilité structurale de cette partie du génome de vBYDV. Ainsi la coupure des brins multimériques se produirait seulement lors de la synthèse de l'ARN. Les connaissances acquises sur PLMVd et vBYDV suggèrent l'existence de mécanismes d'inhibition post-tanscriptionnelle de l'activité des *hammerhead* qui pourrait être essentielle à la formation des conformères circulaires chez les ARN ayant des motifs catalytiques. Les transitions structurales de leur génome semblent être la pierre angulaire de ce type d'inhibition.

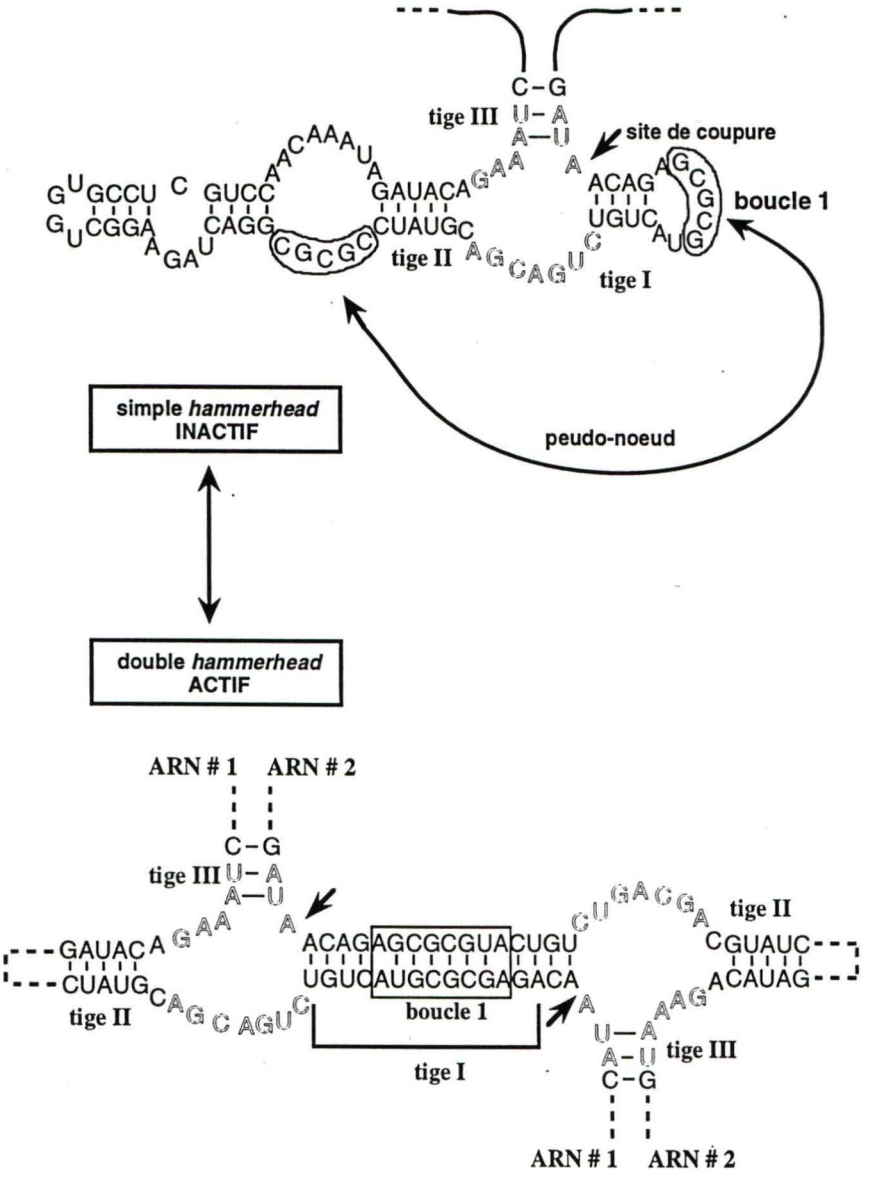


Figure 7. Modulation structurale de l'autocoupure chez vBYDV. La formation d'un pseudo-noeud entre la boucle 1 et la région simple brin de la tige II inhibe l'autocoupure en simple *hammerhead*. La formation d'une structure en double *hammerhead* empêche cette interaction et relève ainsi l'inhibition.

Une meilleure compréhension des mécanismes responsables de la circularisation laisse cependant entrevoir l'existence d'une stratégie bien différente. L'étape de la ligation pourrait en effet assurer l'intégrité des ARN circulaires par deux mécanismes distincts qui dépendent du motif catalytique impliqué dans la réaction d'autocoupure (fig. 8). Chez les ARN satellites sTobRV, sARMV et sCYMV, le motif en épingle à cheveux (hairpin) assurerait un équilibre entre les conformères linéaires et circulaires en raison de la réversibilité de la réaction d'autocoupure *in vitro* (Bruening et al., 1991). L'équilibre thermodynamique de la réaction contribue donc à une constante production de la forme circulaire. L'autoligation est considérée comme étant responsable de la circularisation de ces ARN satellites *in vivo*. La réaction d'autocoupure catalysée par le *hammerhead* est quant à elle irréversible (Hertel et Ulhenbeck, 1995). Cependant, notre groupe a montré que PLMVd est capable d'être circularisé *in vitro* en absence de protéines (Lafontaine et al., 1995). Cette réaction d'autoligation permet de joindre les extrémités 5'-hydroxyle et 2',3'-phosphocycliques résultant de la réaction d'autocoupure des ARN multimériques. L'ARN circulaire contient une liaison de type 2',5' au site de coupure (ou site de ligation) au lieu d'une liaison 3',5' conventionnelle (Côté et Perreault, 1997). Cette modification inactive le *hammerhead* et empêche la réaction d'autocoupure. Il ne semble donc plus nécessaire d'éviter l'action des motifs catalytiques à l'intérieur des ARN circulaires puisque l'autoligation rendrait le site de coupure inapte à promouvoir une réaction d'autocoupure supplémentaire. La localisation des *hammerhead* à l'intérieur de longues hélices pourrait alors favoriser la juxtaposition des deux extrémités. La proximité des deux groupements chimiques dans une hélice d'ARN est en effet suffisante pour permettre l'autoligation (Côté et Perreault, 1997). Cette activité est donc indépendante de la structure active du *hammerhead*. Des expériences sont actuellement en cours afin de vérifier l'existence de la réaction d'autoligation *in vivo* et sa contribution à la formation des ARN circulaires. Les résultats préliminaires suggèrent que la circularisation de PLMVd soit en effet le fruit de l'autoligation (Côté et Perreault, résultats non publiés). Cette possibilité remet donc en cause

la nécessité d'une inhibition post-transcriptionnelle de l'autocoupure par l'adoption de structures alternatives inhibitrices. Les caractéristiques inhérentes au mécanisme de ligation seraient suffisantes pour assurer la stabilité des ARN circulaires. Cependant, la communauté scientifique favorise généralement le modèle de circularisation enzymatique. Des systèmes de ligation *in vitro* à partir de germe de blé permettent en effet la circularisation de certains viroïdes (e.g. PSTVd et PLMVd) (Branch et al., 1982; Beaudry et al., 1995) et ARN satellites (e.g. sTobRV) (Chay et al., 1997). L'enzyme utilise les extrémités 5'-hydroxyle et 2',3'-phosphocyclique et forme une liaison 3',5' conventionnelle. Le nucléotide précédant cette liaison diffère cependant des autres ribonucléotides puisqu'il possède un groupement 2'-monophosphate à la place du groupement 2'-hydroxyle (Kikuchi et al., 1982). Le groupement 2'-hydroxyle n'est donc plus disponible pour une attaque nucléophile de la liaison adjacente, ce qui empêche la réaction d'autocoupure médiée par le *hammerhead* ou même par tout autre motif autocatalytique ayant un mécanisme apparenté. La ligation enzymatique serait également à même de protéger la forme circulaire. Son rôle *in vivo* demeure néanmoins controversé. Une meilleure compréhension des mécanismes de ligation opérant *in vitro* et *in vivo* est donc essentielle pour élucider les éléments responsables de la régulation du ratio des conformères linéaires et circulaires présents à l'intérieur des cellules infectées.

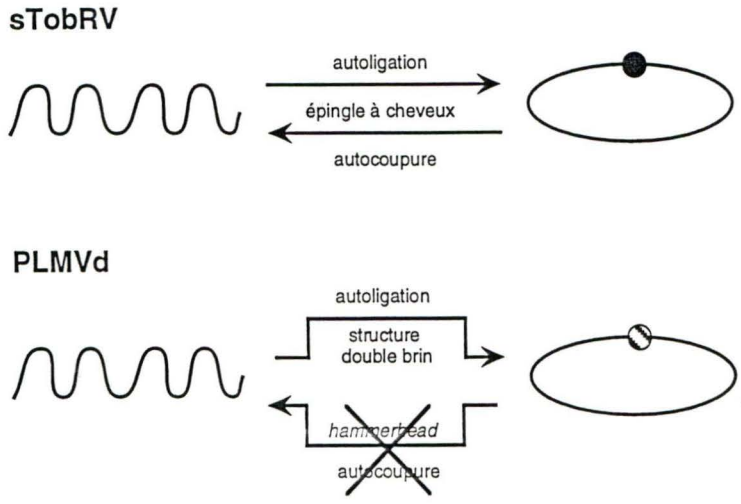


Figure 8. L'autoligation et la protection des ARN circulaires. Le motif en épingle à cheveux catalyse une réaction d'autocoupure réversible qui assure un équilibre entre les conformères linéaires et circulaires (e.g. sTobRV). Chez PLMVd l'autoligation des monomères résultant de la juxtaposition des deux extrémités dans une structure d'ARN double brin mène à la formation d'un ARN circulaire ayant une liaison 2',5' au site de coupure (cercle hachuré). Ce type de liaison empêche la réaction d'autocoupure catalysée par le *hammerhead*. Ces deux mécanismes assurent l'accumulation du conformère circulaire. Le cercle noir représente une liaison 3',5'.

1.2 Le cycle de réplication

La caractérisation des différentes formes accumulées dans les feuilles de pêchers infectés montre que PLMVd se réplique par un mécanisme en cercle roulant symmétrique (fig. 9). Il semble donc que la capacité d'un viroïde à promouvoir la coupure (de façon autocatalytique ou non) de ses intermédiaires multimériques sur une seule polarité (brin (+)), ou bien sur les deux polarités, détermine le type de cercle roulant utilisé. Les viroïdes du groupe B opteraient pour un cycle asymmétrique en raison de leur incapacité à médier la coupure des brins multimériques (-), tandis que l'autocoupure des brins (+) et (-), chez ASBVd et PLMVd, dicterait l'utilisation d'un mécanisme symmétrique (Branch et Robertson, 1984). Le cycle de réplication de PLMVd diffère cependant en ce qui concerne l'activité d'autocoupure des brins multimériques. En effet les résultats suggèrent que la totalité des intermédaires de réplication multimériques soit coupé *in vivo* puisqu'on ne détecte aucun ARN de haut poids moléculaire correspondant à l'une ou l'autre des deux polarités de PLMVd (section 2.1 des Résultats). Une situation semblable a été rapportée pour CChMVd, le dernier viroïde du groupe A caractérisé (Navarro et Flores, 1997). Ces observations contrastent avec les résultats obtenus *in vitro* où le niveau d'autocoupure de ces ARN varie de 50 à 70% (section 1.1 des Résultats; Navarro et Flores, 1997). *In vitro,* un ralentissement de la vitesse de l'ARN polymérase favorise l'adoption transitoire de la structure catalytique des *hammerhead* chez PLMVd (voir fig. 6). L'efficacité de la réaction avoisine alors les 100% (section 1.1 des Résultats). Certaines interactions avec des facteurs cellulaires et/ou la réplication par une ARN polymérase peu processive pourraient donc expliquer le haut niveau d'autocoupure *in vivo*. Alternativement l'absence d'ARN multimériques pourrait être due au fait que PLMVd n'est pas répliqué par un mécanisme en cercle roulant. Cette possibilité s'avère néanmoins peu probable puisque des amplifications par RT-PCR (plus sensibles) démontrent l'existence de formes ayant au moins deux copies de l'unité monomérique.

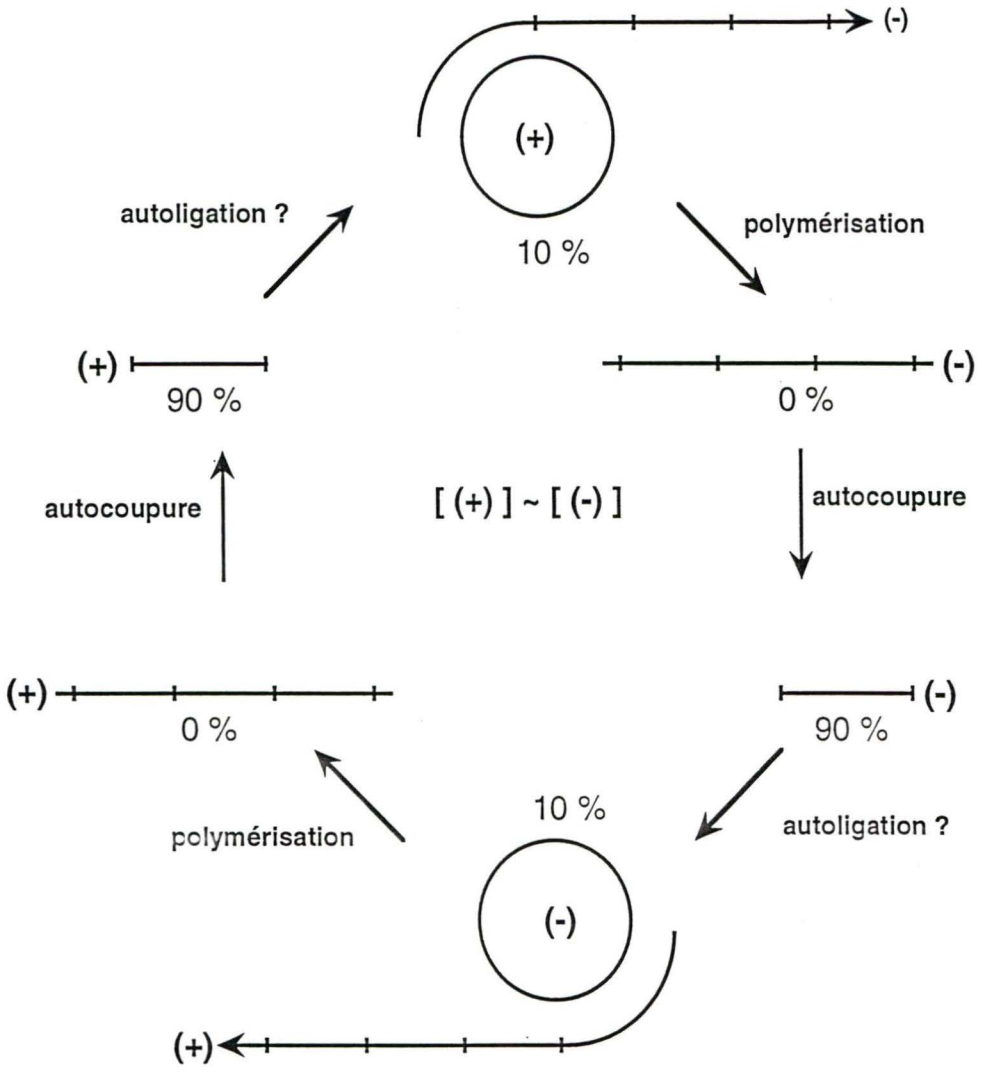


Figure 9. La réplication en cercle roulant chez PLMVd. Représentation des caractéristiques du cercle roulant symmétrique (voir fig. 1 et section 1.1.1 de l'Introduction) de PLMVd *in vivo* . On peut voir les différentes formes ainsi que leur niveau relatif d'accumulation. [ ] indique la concentration de chaque polarité (voir section 2.1 des Résultats).

Une autre particularité du cycle de réplication de PLMVd est qu'il s'accumule principalement sous la forme de monomères linéaires. Ce conformère serait par conséquent très stable *in vivo* (section 2.1 des Résultats). Ces observations vont à l'encontre de la croyance voulant que la circularité des viroïdes soit un gage de leur stabilité. La forme circulaire joue probablement une fonction bien précise dans le cycle vital des viroïdes. Nous croyons que son rôle serait, entre autres, de servir de matrice permettant une réplication en cercle roulant. Cette stratégie évite à la machinerie de réplication de devoir initier la synthèse de l'ARN à l'extrémité 3' du génome répliqué. Ainsi, une initiation interne assure tout de même la production de copies complètes grâce à la synthèse et à la maturation des ARN multimériques. L'importante accumulation du conformère linéaire suggère également que l'étape de ligation est peu efficace. *In vitro,* on sait que PLMVd peut être circularisé par une réaction d'autoligation dont le rendement est d'environ 10% (Côté et Perreault, 1997). Étonnament, ces résultats correspondent au niveau d'accumulation de la forme circulaire *in vivo* (voir section 2.1 des Résultats; fig. 9). Les résultats préliminaires des études de l'autoligation *in vivo* semblent confirmer son rôle dans la recircularisation de PLMVd (Côté et Perreault, résultats non publiés). Le faible ratio circulaire/linéaire chez PLMVd serait donc un effet du mécanisme de ligation peu efficace. Pour le brin (-) de l'ARN satellite sTobRV, ce ratio correspond également au niveau d'autoligation des ARN monomériques (-) *in vitro* (~50%) (Bruening et al., 1991). Il serait actuellement prématuré d'étendre le rôle de l'autoligation dans la formation des conformères circulaires aux autres ARN qui possèdent des motifs catalytiques. Il est cependant intéressant de noter que les extrémités 5'-hydroxyle et 2',3'-phosphocyclique produites suite à une réaction d'autocoupure par l'un ou l'autre des trois motifs catalytiques connus chez ces ARN (*hammerhead*, épingle à cheveux et *delta*) permettent une ligation en absence de protéines *in vitro* (Lafontaine et al., 1995; van Tol et al., 1991; Sharmeen et al., 1989). Le mécanisme commun d'autocoupure résultant de l'attaque du lien phosphodiester par le groupement 2'-hydroxyle adjacent pourrait avoir été mis à profit afin de promouvoir une circularisation indépendante des facteurs cellulaires.

La caractérisation du cercle roulant de PLMVd par des études d'hybridation de type «Northern» a également permis la détection d'un ARN d'environ 1000 nt qui est présent à la fois dans les extraits d'ARN de plants infectés et de plants sains (voir section 2.1 des Résultats, fig. 3A). Les conditions stringentes d'hybridation (60°C dans 50% formamide) laissent croire à d'importantes homologies de séquence entre l'ARN de 1000 nt et le génome de PLMVd. Sa présence chez les plants non infectés indique qu'il s'agit fort probablement d'un ARN cellulaire. N'ayant pu détecter la présence de PLMVd sous forme d'ADN, il apparaît peu probable que cet ARN puisse provenir de la transcription de tels éléments. L'interaction entre PLMVd et l'ARN de 1000 nt au cours de l'infection pourrait provoquer la dégradation du transcrit cellulaire suite à un effet antisens. Sa caractérisation pourrait donc mettre à jour l'existence d'un élément important impliqué dans les mécanismes de pathogénécité des viroïdes. Des observations intéressantes émanent aussi des expériences d'hyrbidation menées sur des échantillons d'ARN préalablement séparés sur gels de polyacrylamide. Cette approche permet de séparer les formes circulaires et linéaires du viroïde, en plus d'apporter une meilleure résolution des fragments d'ARN de faible poids moléculaire. Dans les plants infectés on détecte principalement la forme monomérique linéaire, mais également une série d'ARN plus courts (voir section 2.1 des Résultats, fig. 3D et 3E). On croit que ces ARN seraient des formes tronquées de PLMVd puisqu'on ne les retrouve que dans des extraits d'ARN de feuilles infectées par le viroïde. Leur patron d'accumulation est reproductible et se caractérise par la présence de bandes bien définies. Il serait donc surprenant qu'il résulte d'une dégradation des échantillons d'ARN. La présence, chez PLMVd, de sites particulièrement sensibles à l'hydrolyse ou aux ribonucléases pourrait être à l'origine de la production de ces fragments. Alternativement, ces formes tronquées pourraient être produites lors de la réplication. En effet s'il existe un(des) site(s) de reconnaissance du viroïde circulaire par l'ARN polymérase, il(s) risque(nt) d'être également présent à l'intérieur des ARN linéaires. La prépondérance de la forme monomérique linéaire

provoquerait ainsi la production importante de produits de réplication avortée correspondant aux fragments d'ARN détectés. Mais peu importe l'origine de ces ARN, leur haut niveau d'accumulation suggère leur implication dans le cycle infectieux de PLMVd. Ils pourraient être par exemple des médiateurs de la pathogénécité par l'un ou l'autre des effets antisens ou de mimétisme moléculaire proposés pour provoquer les désordres cellulaires inhérents aux maladies induites par les viroïdes. Une autre hypothèse veut que les motifs *hammerhead* des viroïdes du groupe A soit impliqués dans la coupure (en *trans*) d'ARNm cellulaires (Côté et Perreault, 1997). La régulation négative et spécifique de l'expression de certains gènes engendrerait le processus d'induction des maladies. Notre laboratoire poursuit actuellement des travaux visant à évaluer la validité de cette hypothèse. Les résultats *in vitro* montrent le faible pouvoir catalytique de la copie monomérique complète du viroïde. PLMVd est donc un mauvais ribozyme puisqu'il ne permet pas de catalyser la coupure d'un court substrat en *trans*. Cependant certaines formes tronquées s'avèrent être d'excellents ribozymes (Côté et Perreault, résultats non publiés). Leur production *in vivo* pourrait ainsi contribuer à pomouvoir la coupure d'autres ARN.

1.3 La localisation cellulaire

Les études d'hybridation *in situ* montrent que PLMVd, tout comme ASBVd, se retrouve à l'intérieur des chloroplastes des cellules infectées (Lima et al., 1994) (section 2.1 des Résultats). La localisation cellulaire semble être une caractéristique fondamentale qui différencie les viroïdes du groupe A et du groupe B. Les viroïdes du groupe B sont des ARN nucléaires répliqués par l'ARN pol II en utilisant un mécanisme en cercle roulant asymmétrique résultant de la coupure exclusive des brins multimériques (+). Ceux du groupe A seraient quant à eux répliqués par un cercle roulant symmétrique en raison de leur faculté à catalyser l'autocoupure des intermédiaires multimériques (+) et (-). La réplication se déroule probablement à l'intérieur des chloroplastes, mais on ignore l'identité de l'ARN

polymérase qui en est responsable. Le génome chloroplastique est transcrit par deux systèmes enzymatiques (revue par Link, 1996). Le premier serait une ARN polymérase monomérique (110 kDa) encodée dans le noyau et apparentée aux enzymes de bactériophages (e.g. T7). Son rôle général serait l'expression des gènes d'entretien qui assurent les activités métaboliques de base du chloroplaste (*housekeeping genes*, e.g. ARNr). Le deuxième est une enzyme multimérique (7-13 sous-unités) homologue aux ARN polymérases bactériennes, qui est en partie encodée par le génome chloroplastique. Celle-ci assurerait la transcription des gènes spécifiques à l'activité photosynthétique. La purification biochimique de ces deux activités transcriptionnelles a été réalisée chez plusieurs espèces végétales. Il serait donc intéressant de mettre au point des systèmes transcriptionnels *in vitro* à partir de feuilles de pêchers afin de vérifier si l'un ou l'autre est capable de soutenir la réplication de PLMVd. Les essais préliminaires pour la purification des chloroplastes donnent cependant des résultats mitigés (Bussière et Perreault, résultats non publiés). Les rendements excessivement faibles ne permettent pas, pour le moment, d'envisager la purification des activités transcriptionnelles.

1.4 La structure secondaire

On a identifié plusieurs éléments des structures primaire et secondaire des viroïdes du groupe B qui sont conservés et qui seraient impliqués dans les différentes étapes de leur cycle infectieux (voir section 1.3.1 de l'Introduction et la fig. 2) (revue par Flores et al., 1997). Mis à part les motifs autocatalytiques dont la séquence, la structure et le rôle sont bien connus chez les viroïdes du groupe A, aucun autre élément structural conservé n'a jusqu'à maintenant été identifié. La caractérisation de la structure secondaire *in vitro* de PLMVd révèle pour la première fois l'existence d'éléments potentiellement conservés entre PLMVd et CChMVd, un viroïde du groupe A récemment identifié (Navarro et Flores, 1997). Les motifs catalytiques sont situés dans une région homologue et la majorité des

hélices que l'on retrouve autour de la partie centrale du génome de PLMVd semblent posséder un équivalent chez CChMVd (P1-P4 et P10-P11) (section 3.1 des Résultats). Ces viroïdes sont les seuls à être insolubles dans du LiCl 2M, ce qui supporte l'idée d'une certaine similarité au niveau du repliement global de l'ARN. Nous avons également identifié pour la première fois l'existence d'un pseudo-noeud dans la structure native d'un viroïde. Chez CChMVd une structure homologue pourrait exister dans la région séparant les éléments P4 et P10. Des études d'infectivité du mutant de PLMVd inapte à former le pseudo-noeud sont en cours afin d'établir son importance et son rôle *in vivo*. En dépit de l'absence d'homologie de séquence entre PLMVd et CChMVd, ces deux viroïdes apparaissent donc comme étant très apparentés. De plus, ils possèdent un cercle roulant symmétrique très similaire et accumulent seulement la forme monomérique dont une grande proportion est linéaire. PLMVd et CChMVd apparaissent donc comme des modèles complémentaires idéaux pour poursuivre des études sur la biologie des viroïdes du groupe A, et plus particulièrement des études structure/fonction de ces ARN. Une meilleure connaissance de la structure de CChMVd permettra, par exemple, de pouvoir envisager la construction d'ARN chimériques en remplaçant une région de PLMVd par une structure homologue de CChMVd. Ce genre d'approche s'est avéré particulièrement efficace pour attribuer des rôles biologiques à certaines régions du génome des viroïdes du groupe B (voir section 1.3.1 de l'Introduction). De plus, les propriétés du plant hôte de CChMVd (chrysanthèmes) pourraient peut-être faciliter certaines approches expérimentales (De La Pena et al., 1999). Par exemple la culture en serre des chrysanthèmes est probablement plus facile et plus rapide que ne l'est celle des pêchers. De plus les chrysanthèmes pourraient procurer un matériel biologique plus adéquat pour la purification des chloroplastes.

2. ÉVOLUTION DES VIROÏDES ET DES ARN APPARENTÉS

Les viroïdes, certains ARN satellites de virus de plantes et le VDH forment la famille des ARN circulaires infectieux (Branch et al., 1990). Ces ARN seraient reliés au niveau évolutif (voir Introduction) (Elena et al., 1991). Les transcrits d'ADN répétitif qui possèdent des motifs *hammerhead* (Smaα, rNS2T) pourraient également être apparentés aux viroïdes. Nous avons répertorié toutes les séquences de ces ARN qui étaient disponibles, soit dans des banques de données (NCBI, *National center for biotechnology information*) ou dans des articles scientifiques. La compilation est disponible sur le site http://www.callisto.si.usherb.ca/~jpperra afin de faciliter son utilisation par les chercheurs oeuvrant dans le domaine. Le nombre grandissant de visiteurs et les commentaires reçus confirment l'intérêt engendré par une telle initiative. Nous avons également profité de cette occasion pour proposer une nouvelle nomenclature afin de simplifier l'identification des séquences de viroïdes. Suite à cette compilation, nous avons débuté des études phylogénétiques des viroïdes et ARN apparentés. La grande divergence du génome de ces ARN rend cependant difficile les reconstructions phylogénétiques. La robustesse des arbres obtenus peut être mise en doute même lorsque l'on considère des espèces voisines comme les viroïdes du groupe B. En effet, bien qu'ils partagent plusieurs éléments structuraux, ces espèces font preuve d'une importante divergence au niveau de leur séquence primaire, ce qui rend difficile un alignement non équivoque des positions homologues (voir section 5.1 des Résultats).

Le motif catalytique *hammerhead*, qui est responsable de l'autocoupure des ARN multimériques, est présent chez tous les viroïdes du groupe A et ARN apparentés (excepté le VDH). Nous avons donc entrepris une étude plus approfondie des caractéristiques de ces régions afin de vérifier si elles pouvaient être considérées comme des caractères homologues ayant été conservés au cours de l'évolution. Le *hammerhead* pourrait ainsi permettre d'établir les relations évolutives qui relient ces ARN. Les résultats ont été présentés dans le chapitre 5 (section Résultats) et les paragraphes qui suivent discutent brièvement des hypothèses émises.

Le *hammerhead* est décrits à la section 2 de l'Introduction. La comparaison des *hammerhead* 'naturels' montre une importante conservation de la séquence de la tige/boucle I et plus particulièrement de la tige/boucle II. Ces observations sont surprenantes puisque les variations de la séquence de ces régions n'ont que peu d'influence sur l'activité catalytique *in vitro* (voir section 2 de l'Introduction) L'identité de certains nucléotides pourrait néanmoins influencer le niveau ainsi que la régulation de l'activité d'autocoupure *in vivo*. Le résultat de ces contraintes structurales serait un ralentissement de la vitesse d'évolution des séquences qui composent les *hammerhead*. Ceci permet de retracer des positions homologues à l'intérieur de génomes extrêmement divergents. PLMVd est le seul ARN circulaire pour lequel on dénombre plusieurs variants dont les différences affectent les séquences du *hammerhead*. On peut donc étudier leur changement à une échelle temporelle réduite et vérifier ainsi l'apparente 'inertie' évolutive des tiges/boucle I et II. L'exemple de PLMVd confirme les observations faites en comparant les *hammerhead* des différentes espèces d'ARN. En effet, chez PLMVd la tige III est extrêmement variable tandis que les séquences des tiges/boucle I et II demeurent très stables (Pelchat et al. 1999). De plus, on dénote une synergie potentielle entre la boucle 2 et la tige II qui pourrait contribuer à diminuer le taux de mutations à l'intérieur de cette région. La dernière pb de cette tige

semble affecter l'identité de certains nt de la boucle 2 du *hammerhead* (-) (fig. 10). Une pb $U_{10.5}$-$A_{11.5}$ ou $C_{10.5}$-$G_{11.5}$ coïncide toujours avec les bases respectives G ou A à la position 4 de la boucle 2 (L2.4). Un mésappariement $C_{10.5}$-$A_{11.5}$ provoque quant à lui l'apparition du $G_{L2.4}$, mais également la mutation du résidu $A_{L2.3}$ en G. Le *hammerhead* (+) possède une séquence identique à celle de l'encadré de la figure 10 dans la presque totalité des variants (58/61). Une simple covariation des appariements de type Watson-Crick caractérise le reste de la tige II. Bien qu'on en ignore le rôle *in vivo*, ces observations montrent bien la stabilité des séquence des tiges I et II. Dans les dernières années, des protéines ayant une action bénéfique sur l'activité du *hammerhead* ont été caractérisées (e.g. hnRNP A1; Bertrand et Rossi, 1994). Certaines protéines cellulaires pourraient interagir de façon spécifique avec le *hammerhead* des viroïdes et autres ARN apparentés pour en influencer l'activité. Les tiges I et II seraient des candidats potentiels comme médiateur de cette spécificité d'interaction.

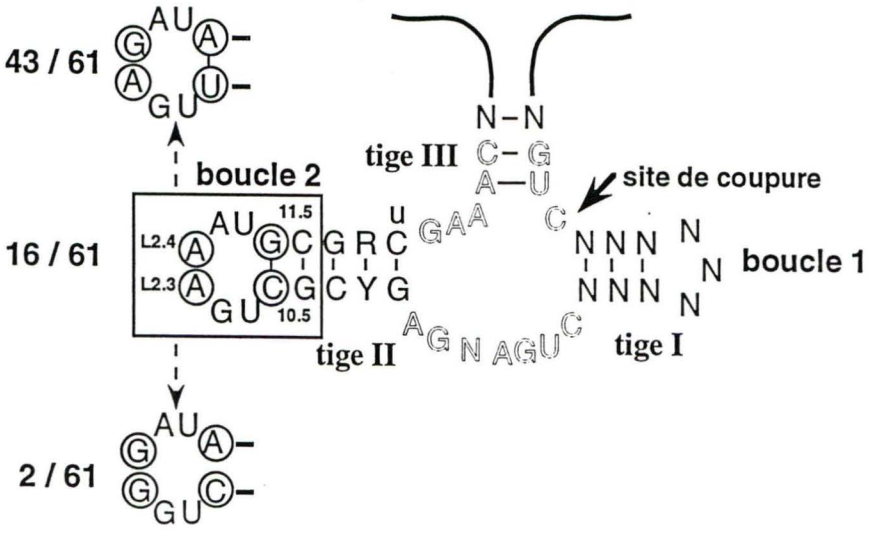
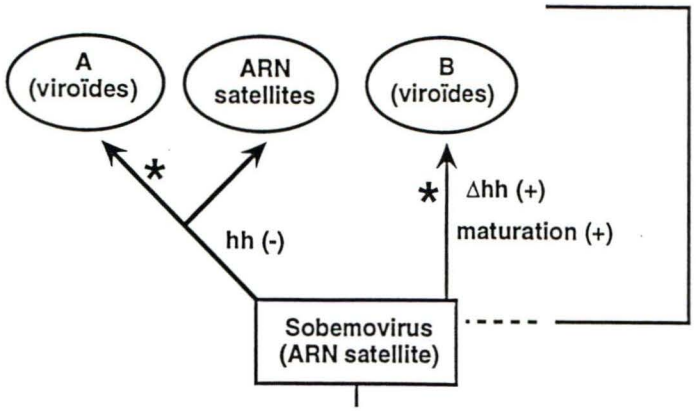


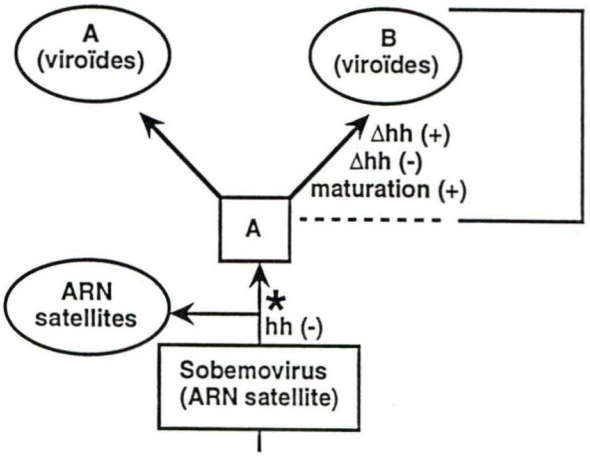
Figure 10. Synergie structurale entre la tige II et la boucle 2 du *hammerhead* (-) chez PLMVd. L'encadré indique la région où l'on observe une covariation des positions 10.5, 11.5, L2.3 et L2.4 (encerclés). Cette région existe sous trois formes chez les 61 variants connus de PLMVd. L'identité de la paire de base 10.5-11.5 semble être reliée à celle des positions L2.3 et L2.4 (voir texte). Les lettres minuscules indiquent des mutations observées chez certains variants.

Les exigences structurales du *hammerhead* en font une région dont la vitesse d'évolution est réduite par rapport au reste du génome. Le *hammerhead* est en quelque sorte l'ARNr des viroïdes et ARN satellites. La conservation des séquences des tiges I et II ainsi que de l'organisation séquentielle des domaines qui composent le *hammerhead* suggère fortement une origine évolutive commune des viroïdes et de certains ARN satellites de virus de plantes. Ce motif catalytique a donc été utilisé comme un caractère homologue afin de reconstruire leur histoire évolutive (voir section 5.1 des Résultats) (fig. 11). Nous proposons qu'un ARN satellite de type sobemovirus, ayant un *hammerhead* sur la polarité (+), soit à l'origine de l'apparition des viroïdes. Suite à la formation d'un nouveau *hammerhead* sur le brin (-) il aurait donné naissance aux viroïdes du groupe A (ainsi qu'à certains ARN satellites). La perte du *hammerhead* (+) et la création d'une nouvelle stratégie pour la maturation des brins multimériques (+) aurait quant à elle mené à l'apparition des viroïdes du groupe B (fig. 11A). Selon cette hypothèse, les viroïdes formeraient un groupe polyphylétique comprenant à la fois des viroïdes et des ARN satellites. Ce qui distingue ces deux types d'ARN est leur capacité à être répliqué ou non par un système enzymatique de la cellule hôte. L'autonomie réplicationnelle des viroïdes serait donc l'effet d'une évolution convergente ayant mené certains ARN satellites à devenir indépendant du virus associé et ce, dans deux lignées évolutives distinctes.

Alternativement, les viroïdes du groupe B pourraient être des descendants directs d'un viroïde apparenté au groupe A (fig. 11B). Leur divergence aurait mené à la perte des motifs *hammerhead* sur les deux polarités ainsi qu'à l'acquisition d'un nouveau mode de maturation des brins multimériques (+). Cette hypothèse rejoint celle du groupe de Diener qui propose que les viroïdes forment un groupe monophylétique (Elena et al., 1991). Les membres du groupe A sont considérés comme un lien évolutif entre les ARN satellites et le reste des viroïdes. En raison de l'absence d'homologies structurales entre ces deux familles de viroïdes il est cependant difficile de pouvoir argumenter en faveur de l'un ou l'autre de ces deux modèles. Néanmoins, la formation des deux groupes de viroïdes est probablement le fruit d'une localisation cellulaire différente. Des interactions avec des composantes cellulaires, soit chloroplastiques ou nucléaires, auraient mené à une importante divergence de ces ARN.



A. Regroupement polyphylétique



B. Regroupement monophylétique

Figure 11. L'évolution des viroïdes à partir d'une forme ancestral d'un ARN satellite de type Sobemovirus. L'hypothèse A suggère que les viroïdes forment un groupe polyphylétique avec certains ARN satellites de virus de plantes. Dans le modèle illustré en B, les viroïdes forment un groupe monophylétique ayant évolué à partir d'un viroïde apparenté au groupe A. Les astérisques représentent l'apparition de la capacité à être répliqué par une ARN polymérase de la cellule hôte. L'acquisition ou la délétion (D) des motifs *hammerhead* (hh) ainsi que l'évolution de stratégies alternatives pour la coupure des brins de polarité plus (+) et moins (-) sont également indiquées. Les espèces encerclées représentent des ARN comtemporains tandis que les formes ancestrales sont encadrées.

3. CONCLUSION

L'objectif des travaux était d'acquérir une meilleure compréhension de la biologie moléculaire des viroïdes du groupe A en utilisant comme modèle PLMVd. La figure 12 compare les caractéristiques principales des viroïdes appartenant aux groupes A et B. Elle résume également les nouvelles connaissances acquises suite aux recherches présentées dans ce travail. Les études menées sur PLMVd ont permis d'identifier de nouvelles caractéristiques structurales en plus de confirmer certaines observations, dont la localisation chloroplastique de ces ARN. De plus, les expériences menées sur la réaction d'autocoupure des motifs catalytiques ont contribué à mieux concevoir son rôle et sa régulation dans le processus de réplication des viroïdes du groupe A et des ARN satellites.

La localisation cellulaire apparaît comme une distinction majeure entre les deux groupes de viroïdes et pourrait expliquer l'absence d'homologie de séquences ou de structures entre ces ARN (fig. 13). La migration à l'intérieur d'organelles différentes a probablement contribuer à la spéciation des viroïdes durant leur évolution. Bien qu'on ignore l'ancêtre de ces agents pathogènes, la comparaison des motifs catalytiques *hammerhead* présents chez les viroïdes du groupe A et les ARN apparentés nous a permis de proposer un schéma évolutif. Cette hypothèse suggère que les viroïdes des deux groupes aient évolué à partir d'un ARN satellite de virus de plantes. La caractérisation de nouveaux ARN ayant des motifs catalytiques est primordiale afin de mieux comprendre l'évolution et l'origine des ces ARN infectieux.

Par leur apparente simplicité, les viroïdes sont des modèles d'étude idéaux en virologie. Les notions provenant de la recherche sur les viroïdes pourraient en effet contribuer à une meilleur compréhension des divers mécanismes utilisés par les virus à ARN pour assurer leur réplication, leur propagation, ainsi que l'induction des maladies. De plus, les viroïdes sont des modèles inestimables pour les études structure/fonction des ARN et le fonctionnement des ARN catalytiques *in vivo*. L'intérêt des viroïdes est donc autant au point de vue de la recherche fondamentale que des aspects plus appliqués, visant à enrayer les pertes considérables en agriculture résultant de l'infection de nombreuses espèces cultivées.

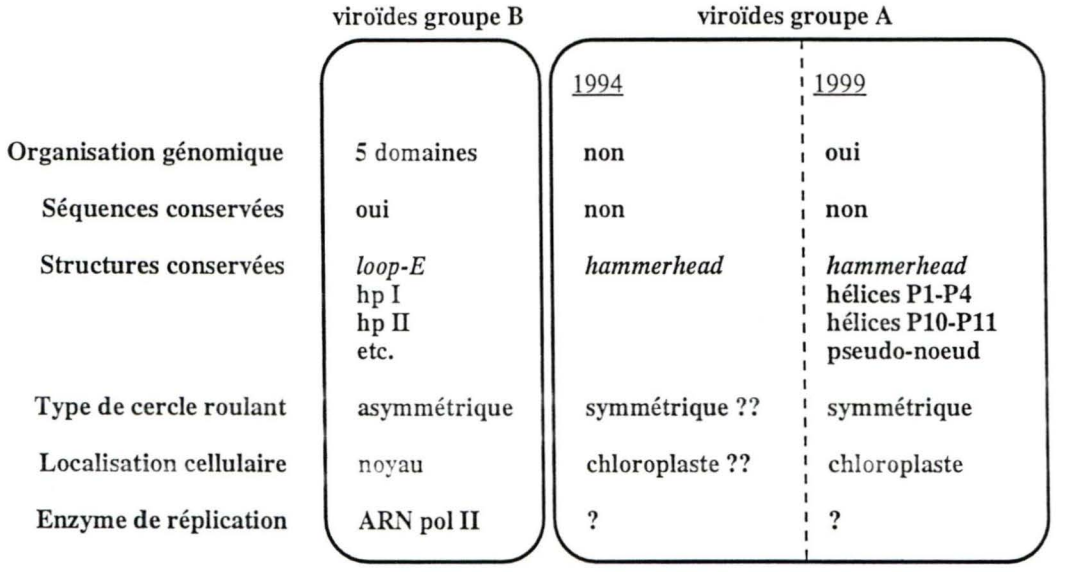


Figure 12. Comparaison des principales caractéristiques des viroïdes des groupes A et B. Depuis le début du projet en 1994, les travaux présentés dans ce travail ont apporté des connaissances nouvelles de la biologie moléculaire des viroïdes du groupe A (1999).

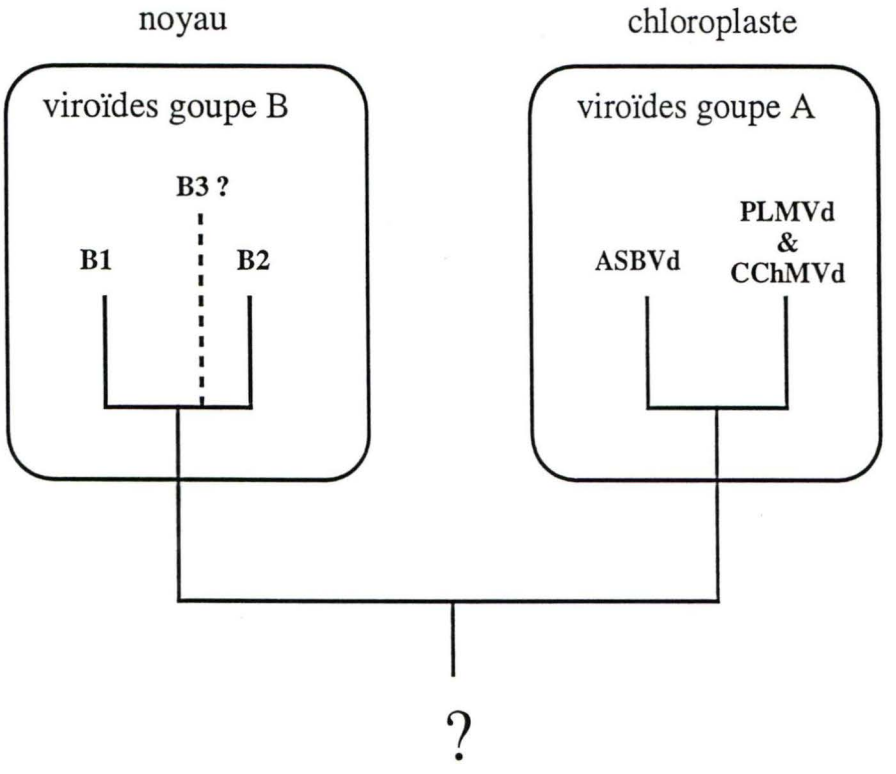


Figure 13. La localisation cellulaire: une distinction majeure entre les deux groupes de viroïdes. Schéma montrant la différence de localisation cellulaire des viroïdes des groupes A et B, ainsi que son rôle potentiel dans la spéciation de ces ARN (le ? représente l'ancêtre commun inconnu).

REMERCIEMENTS

Je tiens à remercier le Dr Jean-Pierre Perreault qui m'a accueilli dans son laboratoire, qui a eu confiance en mes capacités et qui m'a supporté durant toutes ces années. Un merci particulier pour m'avoir permis de relever de nouveaux défis dans ma dernière année. Une année enrichissante et bénéfique qui m'a permis de reprendre véritablement goût à la science. Je voudrais également remercier tous les autres membres du laboratoire et du Département de biochimie avec qui j'ai eu plasir à travailler. Plus particulièrement Daniel Lafontaine, Stéphane Mercure, Fabien Côté, Serge Gravel et Stéphane Veilleux avec lesquels il fut agréable de partager bien des moments qui se sont déroulés bien loin des pipettes et autres objets familiers...Je tiens finalement à remercier Nathalie Coulombe qui m'a épaulé dans les bons et moins bons moments qui ont parsemé la fin du parcours.

RÉFÉRENCES

Ambros, S. and Flores, R. (1998) *In vitro* and *in vivo* self-cleavage of a viroid RNA with a mutation in the hammerhead catalytic pocket. *Nucleic Acids Res*, **26**, 1877-83.

Beaudry, D., Busiere, F., Lareau, F., Lessard, C. and Perreault, J.P. (1995) The RNA of both polarities of the peach latent mosaic viroid self- cleaves *in vitro* solely by single hammerhead structures. *Nucleic Acids Res*, **23**, 745-52.

Bertrand, E.L. and Rossi, J.J. (1994) Facilitation of hammerhead ribozyme catalysis by the nucleocapsid protein of HIV-1 and the heterogeneous nuclear ribonucleoprotein A1. *EMBO J,* **13**, 2904-12.

Berzal-Herranz, A., Joseph, S., Chowrira, B.M., Butcher, S.E. and Burke, J.M. (1993) Essential nucleotide sequences and secondary structure elements of the hairpin ribozyme. *EMBO J,* **12**, 2567-73.

Branch, A.D., Levine, B.J. and Robertson, H.D. (1990) The brotherhood of the small RNA circular pathogens. *Seminars in Virology*, **1**, 143-52.

Branch, A.D., Robertson, H.D., Green, C., Gegenheimer, P., Peebles, C. and Abelson, J. (1982) Cell-free circularization of viroid progeny RNA by RNA ligase from wheat germ. *Science*, **217**, 1147-9

Branch, A.D. and Robertson, H.D. (1984) A replication cycle for viroids and other small infectious RNAs. *Science*, **223**, 450-5.

Brazas, R. and Ganem, D. (1996) A cellular homolog of hepatitis delta antigen: implications for viral replication and evolution. *Science* , **274**, 90-4.

Bruening, G., Gould, A.R., Murphy, P.J. and Symons, R.H. (1982) Oligomers of avocado sunblutch viroid are found in infected avocado leaves. *FEBS Lett*, **148**, 71-8.

Bruening, G., Passmore, B.K., van Tol, H., Buzayan, J.M. and Feldstein, P.A. (1991) Replication of a plant virus satellite RNA: evidence favors transcription of circular templates of both polarities. *Molecular Plant-Microbe Interactions*, **4**, 219-25.

Bussiere, F. and Perreault, J.-P. (1995) On the road to a DNA-protein world. *RNA*, **1**, 451-2.

Bussiere, F., Lafontaine, D. and Perreault, J.P. (1996) Compilation and analysis of viroid and viroid-like RNA sequences. *Nucleic Acids Res*, **24**, 1793-8.

Chay, C.A., Guan, X. and Bruening, G. (1997) Formation of circular satellite tobacco ringspot virus RNA in protoplasts transiently expressing the linear RNA. *Virology*, **239**, 413-25.

Clouet-d'Orval, B. and Ulhenbeck, O.C. (1997) Hammerhead ribozymes with a faster cleavage rate. *Biochemistry*, **36**, 9087-92

Coats, S.R., Zhang, Y, and Epstein, L.M. (1994) Transcription of satellite 2 DNA from newt is driven by a snRNA type of promoter, *Nucleic Acids Res*, **22**, 4697-4704.

Côté, F. and Perreault, J.P. (1997) Peach latent mosaic viroid is locked by a 2',5'-phosphodiester bond produced by *in vitro* self-ligation. *J Mol Biol*, **273**, 533-43.

Daros, J.A. and Flores, R. (1995) Identification of a retroviroid-like element from plants. *Proc Natl Acad Sci USA*, **92**, 6856-60.

Davies, C., Sheldon, C.C. and Symons, R.H. (1991) Alternative hammerhead structures in the self-cleavage of avocado sunbltoch viroid RNAs. *Nucleic Acids Res*, **19**, 1893-8.

De la Pena, M., Navarro, B. and Flores, R. (1999) Mapping the molecular determinant of pathogenecity in a hammerhead viroid: a tetraloop within the *in vitro* branched RNA conformation. *Proc Natl Acad Sci USA*, **96**, 9960-5.

Desvignes, J.C. (1986) Peach latent mosaic and its relation to peach mosaic and peach yellow mosaic virus disease. *Acta Horticulturae*, **235**, 325-32.

Diener, T.O. and Raymer, W.B. (1967) Potato spindle tuber virus: a plant virus with properties of a free nucleic acid. *Science*, **158**, 378-81.

Diener, T.O. (1971) Potato spindle tuber "virus". A replicating, low molecular weight RNA. *Virology*, **45**, 411-28.

Diener, T.O. (1981) Are viroids escaped introns? *Proc Natl Acad Sci USA*, **78**, 5014-15.

Diener, T.O. (1989) Circular RNAs: relics of precellular evolution? *Proc Natl Acad Sci USA*, **86**, 9370-4.

Diener, T.O., Hammond, R.W., Black, T. and Katze, M.G. (1993a) Mechanism of viroid pathogenesis: differential activation of the interferon-induced, double-stranded RNA-activated, M(r) 68,000 protein kinase by viroid strains of varying pathogenicity. *Biochimie*, **75**, 533-8.

Diener, T.O., Owens, R.A. and Hammond, R.W. (1993b) Viroids: the smallest and simplest agents of infectious disease. How do they make plants sick? *Intervirology*, **35**, 186-95.

Ding, B., Kwon, M.O., Hammond, R. and Owens, R. (1997) Cell-to-cell movement of potato spindle tuber viroid. *Plant J*, **12**, 931-6.

Dinter-Gottlieb, G. (1986) Viroids and virusoids are related to group I introns. *Proc Natl Acad Sci USA*, **83**, 6250-4.

Elena, S.F., Dopazo, J., Flores, R., Diener, T.O. and Moya, A. (1991) Phylogeny of viroids, viroidlike satellite RNAs, and the viroidlike domain of hepatitis delta virus RNA. *Proc Natl Acad Sci USA*, **88**, 5631-4.

Epstein, L.M. and Gall, J.G. (1987) Self-cleaving transcripts of satellite DNA from the newt, *Cell*, **48**, 535-43.

Epstein, L.M. and Coats, S.R. (1991) Tissue-specific permutations of self-cleaving newt satellite-2 transcripts, *Gene*, **107**, 213-18.

Epstein, L.M. and Pabon-Pena, L.M. (1991) Alternative modes of self-cleavage by newt satellite 2 transcripts. *Nucleic Acids Res*, **19**, 1699-1705.

Ferbeyre, G., Smith, J.M. and Cedergren, R. (1998) Schistosome satellite DNA encodes active hammerhead ribozymes. *Mol Cell Biol*, **18**, 3880-8.

Flores, R., Hernandez, C., Desvignes, J.C. and Llacer, G. (1990) Some properties of the viroid inducing peach latent mosaic disease. *Res Virol*, **141**, 109-18.

Flores, R., Di Serio, F. and Hernandez, C. (1997) Viroids: The noncoding genomes. *Seminars in Virology*, **8**, 65-73.

Forster, A.C., Davies, C., Sheldon, C.C., Jeffries, A.C. and Symons, R.H. (1988) Self-cleaving viroid and newt RNAs may only be active as dimers. *Nature*, **334**, 265-7.

Forster, A.C. and Symons, R.H. (1987) Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. *Cell*, **49**, 211-20.

Haas, B., Klanner, A., Ramm, K. and Sanger, H.L. (1988) The 7S RNA from tomato leaf tissue resembles a signal recognition particle RNA and exhibits a remarkable sequence complementarity to viroids. *EMBO J*, **7**, 4063-74.

Hammond, R.W., Diener, T.O. and Owens, R.A. (1989) Infectivity of chimeric viroid transcripts reveals the presence of alternative processing sites in potato spindle tuber viroid. *Virology*, **170**, 486-95.

Harders, J., Lukacs, N., Robert-Nicoud, M., Jovin, J.M. and Riesner, D. (1989) Imaging of viroids in nuclei from tomato leaf tissue by in situ hybridization and confocal laser scanning microscopy. *EMBOJ*, **8**, 3941-9.

Haseloff, J., Mohamed, N.A. and Symons, R.H. (1982) Viroid RNAs of the cadang-cadang disease of coconuts. *Nature*, **299**, 316-32.

Hernandez, C. and Flores, R. (1992) Plus and minus RNAs of peach latent mosaic viroid self-cleave *in vitro* via hammerhead structures. *Proc Natl Acad Sci USA*, **89**, 3711-5.

Hertel, K.J., Pardi, A., Uhlenbeck, O.C., Koizumi M., Ohtsuka, E., Uesugi, S., Cedergren, R., Eckstein, F., Gerlach, W.L., Hodgson, R. and Symons, R.H. (1992) Numbering system for the hammerhead. *Nucleic Acids Res,* **20,** 3252.

Hertel, K.J. and Ulhenbeck, O.C. (1995) The internal equilibrum of the hammerhead ribozyme reaction. *Biochemistry*, **34**, 1744-49.

Hiddinga, H.J., Crum, C.J., Hu, J. and Roth, D.A. (1988) Viroid-induced phosphorylation of a host protein related to a dsRNA- dependent protein kinase. *Science*, **241**, 451-3.

Hutchins, C.J., Rathjen, P.D., Forster, A.C. and Symons, R.H. (1986) Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res*, **14**, 3627-40.

Joyce, G.F. (1989) RNA evolution and the origins of life. *Nature*, **338**, 217-24

Keese, P. and Symons, R.H. (1985) Domains in viroids: evidence of intermolecular RNA rearrangements and their contribution to viroid evolution. *Proc Natl Acad Sci USA*, **82**, 4582-6.

Kikuchi, Y., Tye, K., Filipowicz, W., Sänger, H. and Gross, H.J. (1982) Circularization of linear viroid RNA via 2'-phosphomonoester, 3'-5'-phosphodiester bonds by a novel type of RNA ligase from wheat germ and *Chlamydomonas*. *Nucleic Acids Res*, **10**, 7521-29.

Kiss, T., Posfai, J. and Solymosy, F. (1983) Sequence homology between potato spindle tuber viroid and U3B snRNA. *FEBS Lett*, **163**, 217-20.

Kore, A.R., Vaish, N.K., Kutzke, U. and Eckstein, F. (1998) Sequence specificity of the hammerhead ribozyme revisited; the NHH rule. *Nucleic Acids Res*, **26**, 4116-20.

Lafontaine, D., Beaudry, D., Marquis, P. and Perreault, J.-P. (1995) Intra- and intermolecular nonenzymatic ligations occur within transcripts derived from the peach latent mosaic viroid. *Virology*, **212**, 705-9.

Leontis, N.B. and Westhof, E. (1998) A common motif organizes the structure of multi-helix loops in 16s and 23s ribosomal RNAs, *J Mol Biol*, **283**, 571-83.

Lima, M.I., Fonseca, M.E., Flores, R. and Kitajima, E.W. (1994) Detection of avocado sunblotch viroid in chloroplasts of avocado leaves by *in situ* hybridization. *Arch Virol*, **138**, 385-90.

Link,G. (1996) Green life: control of chloroplast gene transcription. *Bioessay*, **18**, 465-71.

Liu, Y.H. and Symons, R.H. (1998) Specific RNA self-cleavage in coconut cadang cadang viroid: potential for a role in rolling circle replication. *RNA*, **4**, 418-29.

Marcos, J.F. and Flores, R. (1992) Characterization of RNAs specific to avocado sunblotch viroid shynthesized *in vitro* by a cell-free system from infected avocado leaves. *Virology*, **186**, 481-8.

Mercure, S., Lafontaine, D., Roy, G. et Perreault, J.-P. (1997) Le motif catalytique d'ARN du virus delta de l'hépatite humaine. *Médecine/Science*, **13**, 662-9.

Miller, W.A. and Silver, S.L. (1991) Alternative tertiairy structure attenuates self-cleavage of the ribozyme in the satellite RNA of barley yellow dwarf virus. *Nucleic Acids Res*, **19**, 5313-20.

Mühlbach, H.P. and Sänger, H.L. (1979) Viroid replication is inhibited by alpha-amanitin. *Nature*, **278**, 185-8.

Navarro, B. and Flores, R. (1997) Chrysanthemum chlorotic viroid: unusual structural properties of a subgroup of self-cleaving viroids with hammerhead ribozymes. *Proc Natl Acad Sci USA*, **94**, 11262-7.

Orgel, L. (1994) The origin of life on the earth. *Scientific American*, **271**, 76-83.

Paul, C.P., Levine, B.J. and Branch, A.D. (1992) Transcripts of the viroid central conserved region contain tertiary structural element found in full-length viroid. *FEBS Lett*, **305**, 9-14.

Pelchat, M., Laurandeau, S., Lévesque, S., Lévesque, D., Lehoux, J., Thompson, D.A., Skrzeczkowski, L.J. and Perreault, J.-P. (1999) Sequencing of peach latent mosaic viroid variants from nine peach cultivars supports a complex secondary structure that include three putative pseudoknots. soumis à *J Virol.*

Querci, M., Owens, R.A., Bartolini, I., Lazarte, V. and Salazar, L.F. (1997) Evidence for heterologous encapsidation of potato spindle tuber viroid in particles of potato leafroll virus. *J Gen Virol*, **78**, 1207-11.

Roossinck, M.J., Sleat, D. and Palukaitis, P. (1992) Satellite RNAs of plant viruses: structures and biological effects. *Microbiological Reviews,* **56**, 265-79.

Sano, T., Candresse, T., Hammond, R.W., Diener, T.O. and Owens, R.A. (1992) Identification of multiple structural domains regulating viroid pathogenicity. *Proc Natl Acad Sci USA*, **89**, 10104-8.

Sano, T. and Ishiguro, A. (1998) Viability and pathogenicity of intersubgroup viroid chimeras suggest possible involvement of the terminal right region in replication. *Virology*, **240**, 238-44.

Semancik, J.S., Tsuruda, D., Zaner, L., Geelen, J.L. and Weathers, L.G. (1976) Exocortis disease: subcellular distribution of pathogenic (viroid) RNA. *Virology*, **69**, 669-76.

Sharmeen, L., Kuo, M.Y.-P. and taylor, J. (1989) Self-ligating RNA sequences on the antigenome of human hepatitis delta virus. *J Virol*, **63**, 1428-30.

Sheldon, C.C. and Symons, R.H. (1989) RNA stem stability in the formation of a self-cleaving hammerhead structure. *Nucleic Acids Res*, **17**, 5665-77.

Spieker, R.L., Haas, B., Charng, Y.C., Freimuller, K. and Sanger, H.L. (1990) Primary and secondary structure of a new viroid 'species' (CbVd 1) present in the Coleus blumei cultivar 'Bienvenue'. *Nucleic Acids Res*, **18**, 3998.

Tornero, P., Conejero, V. and Vera, P. (1994) A gene encoding a novel isoform of the PR-1 protein family from tomato is induced upon viroid infection. *Mol Gen Genet*, **243**, 47-53.

Tornero, P., Conejero, V. and Vera, P. (1996) Primary structure and expression of a pathogen-induced protease (PR- P69) in tomato plants: Similarity of functional domains to subtilisin- like endoproteases. *Proc Natl Acad Sci USA*, **93**, 6332-7.

Tuschl, T. and Eckstein, F. (1993) Hammerhead ribozymes: importance of stem-loop II for activity. *Proc Natl Acad Sci USA*, **90**, 6991-4.

Vaish, N.K., Kore, A.R. and Eckstein, F. (1998) Recent developments in the hammerhead ribozyme field. *Nucleic Acids Res*, **26**, 5237-42.

van Tol, H., Buzayan, J.M. and Bruening, G. (1991) Evidence for spontaneous circle formation in the replication of the satellite RNA of tobacco ringspot virus. *Virology*, **180**, 23-30.

Walter, N.G., Hampel, K.J., Brown, K.M. and Burke, J.M. (1998) Tertiary structure formation in the hairpin ribozyme monitored by fluorescence resonance energy transfer. *EMBO J*, **17**, 2378-91.

Wolff, P., Gilz, R., Schumacher, J. and Riesner, D. (1985) Complexes of viroids with histones and other proteins. *Nucleic Acids Res*, **13**, 355-67.

Zhang, Y. and Epstein, L.M. (1996) Cloning and characterization of extended hammerheads from a diverse set of caudate amphibians, *Gene*, **172**, 183-90.

ANNEXES

1.1 ARTICLE:

Evidence for a model ancestral viroid.

Frédéric Bussière, Daniel Lafontaine, Fabien Côté,

Danièle Beaudry et Jean-Pierre Perreault

Département de Biochimie

Université de Sherbrooke

Sherbrooke, Québec, J1H 5N4, Canada

Article publié dans: *Nucleic Acids Symposium Series*, Vol. 33, 143-144, 1995.

1.2 ARTICLE:

On the road to a DNA-protein world.

Frédéric Bussière et Jean-Pierre Perreault

Département de Biochimie

Université de Sherbrooke

Sherbrooke, Québec, J1H 5N4, Canada

Article pubilé dans: *RNA*, Vol.. 1, 451-452, 1995.

Nucleic Acids Symposium Series No. 33 143–144

Evidence for a model ancestral viroid

Frédéric Bussière, Daniel Lafontaine, Fabien Côté, Danièle Beaudry and Jean-Pierre Perreault*
Département de Biochimie, Faculté de Médecine, Université de Sherbrooke, Sherbrooke, Québec J1H 5N4, Canada

ABSTRACT

The generation of a phylogenetic tree of viroids and viroid-like plant satellite RNAs via computer analysis, coupled with several conspicuous biochemical characteristics of the rolling circle replication of these RNAs -including both self-cleavage and self-ligation- leads us to propose the peach latent mosaic viroid (PLMVd) as a current "living fossil" dating from a precellular world. Incorporated within this proposal is a revised mechanism of PLMVd rolling circle replication which requires a minimal protein involvement.

INTRODUCTION

ids are small (246-375 nt) single-stranded circular RNAs infect higher plants causing diseases in numerous crop species which result in significant economic losses in the agricultural industry (1,2). Both viroids and "viroid-like" plant satellite RNAs have been proposed as examples of "living fossils" from a precellular world (see Ref. 3). The fundamental belief of this hypothesis is that these biological macromolecules exhibit both genotypical and phenotypical functions.

Using the numerous viroid sequences published during the last five years, we have refined the phylogenetic analysis of viroids and viroid-like plant satellite RNAs (4). This modification of the phylogenetic tree was achieved using the maximum parsimony algorithms included in the PAUP package (5). This analytical method, coupled with the use of a larger number of species, permitted the prediction of a more parsimonious tree than any published to date (6-8). The resulting tree, which is summarized in figure 1, supports a monophylogenetic origin of these RNAs as proposed previously (3). As illustrated in figure 1, the viroid sequences are clustered in three groups (potato spindle tuber viroid (PSTVd) group, apple scar skin viroid (ASSVd) group and avocado sunblotch viroid (ASBVd) group), while satellite RNAs are clustered together in a fourth group. The cluster formed by the ASBVd, the PLMVd and the carnation stunt associated viroid (CarSAV, which is not yet formally classified as a viroid) supports the existence of the ASBVd group. The members of this group, which appears as an evolutionary link between the classical viroids and the related ite RNAs, have the capacity to self-cleave by nammerhead structures.

Here, we report several experiments designed to identify the member of the ASBVd group that most closely resembles the potential ancestral organism of both classical viroids and satellite RNAs.

PHYLOGENETIC EXPERIMENTS

Two different approaches were used for rooting the proposed tree. Firstly the satellite RNAs were considered as the outgroup taxa; and, secondly, a minimal number of species, including members of each group, were used to computer generate phylogenetic trees by the exhaustive method using various outgroup taxa (including viroid-like domains of either the mitochondria transcripts from satellite 2 DNA or the human delta hepatitus virus, RNA replicon domains, etc). These analyses produced no precise root as no selection of outgroup taxa was perfect. However, all tested outgroups permitted the rooting of the tree in the first part of the ASBVd group branch (fig. 1). These results support the hypothesis that ASBVd group is an evolutionary link between the classical viroids and the satellite RNAs. Furthermore, these results suggest that the hammerhead motif was part of the ancestral viroid world as it is found in the members of ASBVd group.

BIOLOGICAL CHARACTERISTICS

To determine whether a viroid from the ASBVd group may be a current "living fossil", we considered their known

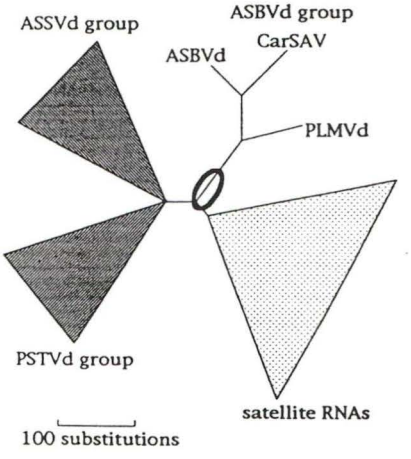


Figure 1.
Schematic representation of the proposed phylogenetic tree of viroids and viroid-like plant satellite RNAs. The triangles are the lineages of the PSTVd group and ASSVd group. The triangle size represents the diversity of these lineages. For ASBVd group, each member is specificaly identified. The circle (bold) defines the region where several taxa rooted the tree.

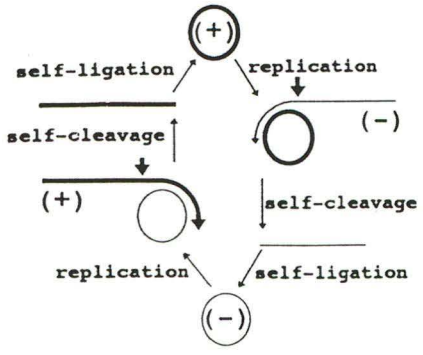


Figure 2.
Simplified rolling circle mechanism proposed for PLMVd (see text for description).

replication characteristics. We characterized the hammerhead self-cleavage of dimeric, monomeric and mutated transcripts derived from RNA of both polarities of PLMVd (9). Several experiments demonstrated that transcripts of both polarities derived from PLMVd self-cleave exclusively by single hammerhead structures. This observation contrasts with the case of ASBVd and CarSAV whose mechanisms involve mostly double hammerhead structures. Single hammerhead cleavage is associated with viroid-like satellite RNAs, and this PLMVd characteristic suggests that it may be the more likely candidate as a current "living fossil" dating from the precellular world.

Furthermore, we have demonstrated the non-enzymatic self-ligation of transcripts both polarities of PLMVd (10), the first such description of this process with viroid sequences. Self-ligation occurs when the 5'-hydroxyl and the 2'-3'-cyclic phosphate termini produced by the hammerhead self-cleavage of the viroid RNA are juxtaposed by the viroid rod-like structure, and a phosphodiester bond is formed between the two following hydrolysis of the cyclic phosphate. The intramolecular self-ligation of PLMVd transcripts could contribute to the production of properly circularized viroids during rolling circle replication. In addition, these results suggest that an ancestral viroid may not require the existence of a specific ligase to produce circularized progeny.

ROLLING CIRCLE MECHANISM

The discovery of both the self-ligation and the self-cleavage behavior of PLMVd transcripts suggests a reworking of the rolling circle mechanism of PLMVd replication (Fig. 2). Initially, the plus polarity circular RNA is copied to yield a RNA of minus polarity. In contrast to classical viroids (like PSTVd), the RNA polymerase II from wheat germ did not support the replication of PLMVd (unpublished data, F. Lareau and J.-P. Perreault) and ASBVd (11), in fact the RNA polymerase responsible for this step remains to be identified. Secondly, once the hammerhead motif is synthesized, the transcript rapidly self-cleaved. This reaction occurs concurrently with the polymerization step, hence no accumulation of long multimeric replicates is expected for PLMVd. Thirdly, the polymerization of the minus polarity RNA is persued, and, when a unit-length transcript is complete, the RNA folds into its stable rod-like structure which brings both termini into the close proximity required for self-ligation. This last reaction may be partly nonenzymatic, enhanced by a protein cofactor or catalyzed by a protein *in vivo*. The self-ligation has the advantage of favorizing the formation of a 2',5'-phosphodiester bond which prevents further intramolecular self-cleavage. The same three steps are repeated to yield the progeny viroids. In contrast to the proposed replication mechanism for classical viroids, which requires different enzymes for each step, only the polymerisation described here requires the involvement of an exogenous enzymatic activity. This simplified rolling circle mechanism could be important in the replication of ancestral viroids because of the minimal requirement for exogenous enzymatic activity.

The sequences required for both the self-cleavage and self-ligation of PLMVd, are localized in the left arm region and concentrated in ~100 nt of this viroid. This region is the replicational domain of PLMVd and could have been shared with an ancestral viroid.

ACKNOWLEDGEMENTS

This work was sponsored by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to J-PP. FL, DL and FC are in receipt of predoctoral fellowships from respectively Fondation George Phoenix, Fonds FRSQ-FCAR and Université de Sherbrooke. J-PP is a Scholar of the Medical Research Council of Canada (MRC).

REFERENCES

1. Symons,R.H. (1990) *Semin. Virol.*, **1**, 75-81.
2. Diener,T.O. (1993) *Trends Microbio.*, **1**, 289-294.
3. Diener,T.O. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9370-9374.
4. Bussière, F., Lafontaine,D. and Perreault,J.-P. (1995) unpublished results.
5. Swofford,D.L. (1992) PAUP-Phylogenetic Analysis Using Parsimony. Version 3.0 s.
6. Elena,S.F., Dopazo,J., Flores,R., Diener,T.O. and Moya,A. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 5631-5634.
7. Hernandez,C. and Flores,R. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 3711-3715.
8. Hernandez,C., Elena,S.F., Moya,A. and Flores,R. (1992) *J. Gen. Virol.*, **73**, 2503-2507.
9. Beaudry,D., Bussière,F., Lareau,F., Lessard,C. and Perreault,J.-P. (1995) *Nucleic Acids Res.*, **23**, 745-752.
10. Lafontaine,D., Marquis,P., Beaudry,D. and Perreault,J.-P. (1995) submitted to *J. Virol.*
11. Lima,M.I., Fonseca,M.E.N., Flores,R. and Kitajima,E.W. (1994) *Arch. Virol.*, **138**, 385-390.

*To whom correspondence should be addressed

RNA (1995), *1*:451-452. Cambridge University Press. Printed in the USA.

LETTER TO THE EDITOR

On the road to a DNA-protein world

FRÉDÉRIC BUSSIÈRE and JEAN-PIERRE PERREAULT

Département de Biochimie, Faculté de Médecine, Université de Sherbrooke, Sherbrooke, Québec J1H 5N4, Canada

Following the RNA world and the appearance of ribonuclease reductase activity, it is proposed that the progression of genetic templates from single-stranded RNA to DNA occurred gradually via an increasing deoxyribonucleotide content in the RNA molecules coupled to an increased polymerase stringency, rather than in one step corresponding to a major polymerase evolution (Lazcano et al., 1992). This hypothesis is supported by both DNA (Lazcano et al., 1992) and RNA polymerase data (Conrad et al., 1995) that indicate that the early polymerases possessed rather broad substrate and template specificities (ribose versus deoxyribose). This hypothesis envisions mixed RNA-DNA molecules (M-molecules) composed of randomly interspersed ribo- and deoxyribonucleotides. The obvious question is whether such M-molecules served exclusively template functions or also served catalytic functions.

With the development of chemical synthesis strategies for RNA compatible with those for DNA, M-molecules can now be prepared (Usman & Cedergren, 1992) and this question properly addressed. Synthetic M-molecules possess several conspicuous characteristics, as illustrated by the following examples. Predominantly deoxyribonucleotide-containing "hammerhead" ribozymes (i.e., nucleozymes) exhibit a significant level of cleavage, showing that ribozyme-type catalysis is not restricted solely to pure RNA (Bratty et al., 1993). Furthermore, DNA molecules with catalytic functions (i.e., deoxyribozymes) have been developed using the powerful in vitro selection approach. Breaker and Joyce (1994) synthesized a deoxyribozyme with the ability to cleave RNA molecules, and Cuenoud and Szostak (1995) developed a deoxyribozyme that catalyzes the ligation of DNA substrates. In addition, several reports have shown that both M-molecules and pure DNA molecules of sequences corresponding to RNA substrates may act as ribozyme substrates (Cech et al., 1992; Bratty et al., 1993; Perreault & Altman, 1992). Moreover, the presence of the deoxyribonucleotides in ribozymes confers greater stability toward both aqueous hydrolysis, which is promoted by the 2'-hydroxyl groups, and ribonuclease degradation (Bratty et al., 1993). Thus, the introduction of several deoxyribonucleotides into RNA molecules does not necessarily result in a severe reduction in either their catalytic properties or their substrate efficiency. Hence, M-molecules appear to fold in a manner analogous to pure RNA. This point is also illustrated by the introduction of a single ribonucleotide into a DNA helix driving the conformation to the A-type helix characteristic of RNA, as observed in three-dimensional studies of some Okazaki fragments (Egli et al., 1993).

The initial incorporation of deoxyribonucleotides into RNA was totally at random due to the indiscriminate nature of the existing polymerases. The different nucleotides within the resulting M-molecules were distinguished solely by their base-pairing properties and not by their sugar moieties. As a result, catalytically inactive products lacking a ribonucleotide at a critical position may have existed; however, whether all M-molecules required a ribonucleotide at a given position is not clear, because mutant sequences could conceivably have released these molecules from the ribonucleotide requirement. Additionally, catalytically inactive products might have been replicated, and their descendants either re-established catalytic properties or became specialized for the carrying of genetic information.

Thus, M-molecules are not only polymerized by enzymes but also possess catalytic activities analogous to those of RNA and may be more stable than pure RNA. Clearly they satisfy all the requirements for early genetic material, and, when coupled with the concurrent evolution of polymerase stringency, provide the likely vehicle for the bridge from the ancestral RNA world to the modern DNA-protein world.

Reprint requests to: Jean-Pierre Perreault, Département de Biochimie, Faculté de Médecine, Université de Sherbrooke, Sherbrooke, Québec J1H 5N4, Canada; e-mail: jp.perre@courrier.usherb.ca.

REFERENCES

Bratty J, Chartrand P, Ferbeyre G, Cedergren RJ. 1993. The hammerhead RNA domain, a model ribozyme. *Biochim Biophys Acta 1216*:345-359.

Breaker RR, Joyce G. 1994. A DNA enzyme that cleaves RNA. *Chem Biol 1*:223-229.

Cech TR, Herschlag D, Piccirilli JA, Pyle AM. 1992. RNA catalysis by a group I ribozyme. *J Biol Chem 267*:17479-17482.

Conrad F, Hanne A, Gaur RK, Krupp G. 1995. Enzymatic synthesis of 2'-modified nucleic acids: Identification of important phosphate and ribose moieties in RNAse P substrates. *Nucleic Acids Res* 23. Forthcoming.

Cuenoud B, Szostak JW. 1995. A DNA metalloenzyme with DNA ligase activity. *Nature 375*:611-614

Egli M, Usman N, Rich A. 1993. Conformational influence of the ribose 2'-hydroxyl group: Crystal structures of DNA-RNA chimeric duplexes. *Biochemistry 32*:3221-3237.

Lazcano A, Valverde V, Hernandez G, Gariglio P, Fox GE, Oro J. 1992. On the early emergence of reverse transcription: Theoretical basis and experimental evidence. *J Mol Evol 35*:524-536.

Perreault JP, Altman S. 1992. Important 2'-hydroxyl groups in model substrate for M1 RNA, the catalytic RNA subunit of RNAse P from *Escherichia coli*. *J Mol Biol 226*:399-409.

Usman N, Cedergren RJ. 1992. Exploiting the chemical synthesis of RNA. *Trends Biochem Sci 17*:334-339.

1.3 **LETTRES D'APPROBATION DES CO-AUTEURS.**